R.S. Guter
L.D. Kudryavtsev
B.M. Levitan

# Elements of The Theory of Functions

*INTERNATIONAL SERIES OF MONOGRAPHS IN*
**PURE AND APPLIED MATHEMATICS**
GENERAL EDITORS: I. N. SNEDDON AND M. STARK
EXECUTIVE EDITORS: J. P. KAHANE, A. P. ROBERTSON AND S. ULAM

VOLUME 90

# ELEMENTS OF THE THEORY OF FUNCTIONS

R. S. GUTER, L. D. KUDRYAVTSEV,
B. M. LEVITAN

# Elements of the Theory of Functions

TRANSLATED BY

H. F. CLEAVES

TRANSLATION EDITED BY

I. N. SNEDDON
*Simson Professor in Mathematics*
*University of Glasgow*

PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
TORONTO · PARIS · BRAUNSCHWEIG

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1
Pergamon Press (Scotland) Ltd., 2 & 3 Teviot Place, Edinburgh 1
Pergamon Press Inc., 44–01 21st Street, Long Island City, New York 11101
Pergamon of Canada, Ltd., 6 Adelaide Street East, Toronto, Ontario
Pergamon Press S.A.R.L. 24 rue des Écoles, Paris 5e
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

Copyright © 1966
Pergamon Press Ltd.

First English edition 1966

This is a translation of the original Russian
Элементы теории функций (Elementiy teorii funktsii),
published in 1963 by Fizmatgiz, Moscow

Library of Congress Catalog Card No. 66–14658

2518/66

# CONTENTS


FOREWORD ix

EDITOR'S PREFACE TO THE RUSSIAN EDITION xi

CHAPTER I. FUNCTIONS OF A REAL VARIABLE (BASIC IDEAS AND THEOREMS) 1

§ 1. Linear Point Sets. Power and Category 1
§ 2. The Lebesgue Measure of Linear Sets 16
§ 3. Basic Classes of Functions 22
§ 4. The Derivative and its Generalization 31
§ 5. The Integration of Functions 36
§ 6. Sequences of Functions 52
§ 7. Orthogonal Systems of Functions 60
§ 8. Functions of Several Variables 70
§ 9. Basic Function Spaces 78

CHAPTER II. THE INTERPOLATION AND APPROXIMATION OF FUNCTIONS 86

§ 1. Introductory Remarks 86

§ 2. The Interpolation of Functions by Polynomials 87
2.1. The Simplest Interpolation Problem. The Lagrange and Newtonian Interpolation Polynomials 87
2.2. Interpolation by Trigonometric Polynomials 92
2.3. General Interpolation Problems. Hermite's formula 95
2.4. The Convergence of Interpolation Polynomials 96
2.5. Operations on Interpolation Polynomials Leading to Convergent Processes 103
2.6. Generalized Interpolation Polynomials 104
2.7. Remarks 105

§ 3. Uniform Approximations of Functions of one Variable by Polynomials and their Generalizations 107
3.1. General Remarks on the Approximation of Functions 107
3.2. Weierstrass' Theorem 108
3.3. Bernstein Polynomials 109
3.4. The Best Uniform Approximations of Functions by Polynomials of a Given Degree 110
3.5. Chebyshev's Results for Best Approximations 112
3.6. The Approximate Formation of Best Approximating Chebyshev Polynomials 114

3.7. The Best Approximations to Continuous and Differentiable Periodic Functions by Trigonometric Polynomials of a Given Degree 118
3.8. The Best Approximations of Continuous and Differentiable Functions by Algebraic Polynomials of a Given Degree 123
3.9. The Best Approximations of Functions by Integral Functions 129
3.10. Weighted Approximations of Functions on the Whole Numerical Axis 131
§ 4. Methods of Uniform Approximation of Functions 132
4.1. The Approximation of Periodic Functions by Fourier Sums 132
4.2. Linear Methods of Approximation of Periodic Functions by Trigonometric Polynomials (the methods of Féjer, LaVallée Poussin and Bernstein–Rogosinski) 135
4.3. Linear Methods of Approximation of Functions by Algebraic Polynomials 139
4.4. The Best Linear Methods of Approximation of Functions 143
§ 5. The Approximation of Functions of one Variable in the Mean 146
5.1. General Remarks 146
5.2. Best Approximations in the Mean 148
5.3. Systems of Functions which Best Approximate Some Class of Functions 151
§ 6. The Approximation of Functions of Many Variables 152
6.1. Basic Ideas 152
6.2. The Theory of Best Approximations of Functions of Many Variables 156
§ 7. The Theory of Approximations in Banach Spaces 158
7.1. General Ideas. Best Approximations in Hilbert Spaces. The Method of Least Squares for the Construction of best Approximations 158
7.2. The Connection Between the Best Approximations and the Entropy of the Sets 160
7.3. The Theory of Best Chebyshev Approximations on Compacta 162
7.4. Some General Remarks on the Theory of Approximation of Functions 164

CHAPTER III. ALMOST-PERIODIC FUNCTIONS 170
§ 1. Uniform Almost-periodic Functions on a Straight Line 170
1.1. Various Definitions of Almost-periodic Functions 170
1.2. The Simplest Basic Properties of Almost-periodic Functions 171
1.3. Fourier Series 172
1.4. Formal Operations on Fourier Series 175
1.5. Basic Theorems 176
1.6. The Approximation Theorem 178
1.7. The Convergence of Fourier Series for Some Classes of Uniform Almost-periodic Functions 179

1.8. The Relation Between the Fourier Exponents and Almost-periods 181
1.9. Kronecker's Theorem 182
1.10. Limited Periodic Functions 183
1.11. The Theorem of the Argument of a Uniform Almost-periodic Function 183
1.12. $N$-almost-periodic Functions 184
1.13. Linear Differential Equations with Almost-periodic Coefficients 186
§ 2. Various Generalizations of Almost-periodic Functions 188
2.1. Introductory Remarks 188
2.2. The Definition and Simplest Properties of Stepanov Almost-periodic Functions 188
2.3. The Definition and Simplest Properties of Weyl Almost-periodic Functions 190
2.4. The Mean Value Theorem for $W^p$-almost-periodic functions. Parseval's Equation for $W^2$- and $S^2$-almost-periodic Functions. The Approximation Theorem 191
2.5. Besicovitch Almost-periodic Functions 192
2.6. The Idea of Almost-periodic Functions on Groups 193
§ 3. Analytic Almost-periodic Functions 195
3.1. The Definition of Analytic Almost-Periodic Functions and Their Simplest Properties 195
3.2. Dirichlet Series 197
3.3. The Convergence of the Dirichlet Series for Analytic Almost-periodic Functions 199
3.4. The Behaviour of Almost-periodic Functions when $\sigma = +\infty$ (analogues of the theorems of Weierstrass–Sokhotskii and Picard theorems) 199
3.5. Harmonic Almost-periodic Functions 200
3.6. The Mean Movement and Density of Values of Analytic Almost-periodic Functions 201

REFERENCES 206

INDEX OF SYMBOLS 209

INDEX 211

LIST OF OTHER TITLES IN THE SERIES 217

# FOREWORD

THIS publication deals with the basic ideas of the theory of functions in a real domain. Its chief contents are questions of the approximation and interpolation of functions of a real variable, and also of the theory of almost-periodic functions. These chapters are preceded by a chapter on the general ideas of the theory of functions of a real variable.

The book is intended for specialists in sciences in which mathematics plays a prominent role and also for mathematicians who are not acquainted with the theory of functions.

# EDITOR'S PREFACE TO THE RUSSIAN EDITION

In order to strictly substantiate the large number of facts accumulated in analysis, at the end of the nineteenth century there arose a new discipline—the theory of sets and functions, which has been developed rapidly of recent years. The development of this theory has significantly broadened and deepened the ideas considered in analysis. At the beginning of the nineteenth century the extremely important idea of the measure of a set (Lebesgue) was put forward, on the basis of which the idea of the Lebesgue integral was created. These two fundamental ideas constitute the basis of the metric theory of functions of a real variable, which deals with the study of the properties of functions, derivatives, integrals and series of functions by means of set measure. Methods and ideas of the theory of functions of a real variable have contributed to the development of a series of new mathematical disciplines and, in addition, have penetrated other regions of mathematics (topology, the theory of probability, functional analysis, the calculus of variations, differential equations, etc.). For this reason the role and significance of the theory of functions in contemporary mathematics is very great.

Chapter I of the present book was written by R. S. Guter and gives an exposition of some basic ideas and theorems from the theory of sets, and the metric theory of functions of a real variable. It can be understood by anyone who has mastered a course of analysis in a technical institute. The contents of this chapter do not reflect the present state of the theory of functions of a real variable, but contain only basic information from this theory. Their proofs can be found in I. P. Natanson's book *The theory of functions of a real variable* (see References, Chapter I [7]). For further detailed treatment of this theory more specialist manuals must be consulted.

Chapter II, "The interpolation and approximation of functions", was written by L. D. Kudryavtsev. To understand it, a knowledge

of the theory of the Lebesgue integral is needed, as given in Chapter I. In addition to purely theoretical interest, at the present time the role of the theory of interpolation and approximation of functions increases more and more in connection with the intensive development of numerical analysis and computer science. It must be noted that in Chapter II the material on the approximation of functions deals with the latest contemporary results in this field. Therefore Chapter II will be useful and interesting both for engineers and for mathematicians who are not specialists in the theory of approximation of functions.

Chapter III was written by B. M. Levitan and gives the basic facts of the theory of almost-periodic functions, which are a natural extension of periodic functions. To read the first section of Chapter III it is sufficient to have mastered a technical college course in analysis. To understand §§ 2 and 3 it is necessary to have an acquaintance with the basic theory of the Lebesgue integral (as given in Chapter I) and the theory of functions of a complex variable in the specialist mathematics courses of technical colleges.

At the end of the book there is a list of references for each chapter in which are given some of the basic monographs relating to the questions under consideration. This list does not in any way pretend to be complete.

As a whole the present work is a suitable textbook for anyone who has mastered the bulk of the mathematics course in a technical college.

P. L. Ul'yanov

CHAPTER I

# FUNCTIONS OF A REAL VARIABLE (BASIC IDEAS AND THEOREMS)

## § 1. LINEAR POINT SETS. POWER AND CATEGORY

**1.** By the word *set* is understood the combination of some finite or infinite number of objects. In a number of cases the word "set" can be replaced by other words.

For instance, in geometry a set of points, possessing some common property, is called a *locus*. Thus a circle is the locus (set) of points of a plane, situated at the same distance from a given point (the centre). The set of all integral curves of a given differential equation is called a *family*. In addition the term *aggregate* is sometimes used. All these words are synonyms of the word "set" and are sometimes used instead of it.

Sets are usually denoted by capital Latin letters, e.g. $M, E, \cdots$, and their elements by small letters, e.g. $x, y, \cdots$ To denote the fact that $x$ is an element of the set $M$, the sign $\in$ is used. The notation $x \in M$ is read as "$x$ belongs to $M$" or "$x$ is an element of $M$", or finally "the element $x$ enters into (is contained in) the set $M$". If $x$ does not belong to $M$, then it is written $x \overline{\in} M$ or $x \notin M$.

The initial subject of investigation will be *linear point sets*, i.e. sets whose *elements* are points of a straight line, or more precisely speaking, points of a numerical axis. The *segment* and the *interval* are the simplest examples of linear point sets.

A *segment* is a set of points $x$ of a numerical axis, which satisfy the inequality $a \leqslant x \leqslant b$, and an *interval* a set of points $x$ of a numerical axis, which satisfy the inequality $a < x < b$.

A segment is usually denoted by the symbol $[a, b]$, and an interval by the symbol $(a, b)$. From the definition it is obvious that the segment $[a, b]$ is different from the interval $(a, b)$ in that the points $x = a$ and $x = b$ belong to the segment but not to the interval.

If each element of the set $B$ is at the same time also an element of the set $A$, then we say that $B$ is a *subset* of $A$, or that $B$ is

*contained in* $A$. In this case write $B \subseteq A$ or $A \supseteq B$. Here also we do not rule out the possibility that the set $B$ may consist of *all* the elements of $A$.

If all the elements of $B$ are also elements of $A$ and, conversely, all elements of $A$ are elements of $B$, we say that the sets $A$ and $B$ *coincide*. This fact is written as the equation $A = B$.

When we wish to stress the fact that $B$ is a subset of $A$, but does not coincide with $A$, we write $B \subset A$ and we say that $B$ is a *proper subset* of $A$ (or a *proper part* of $A$).

The set, which does not contain even one element, is said to be *empty*. Thus the set of real roots of the equation $x^4 + 1 = 0$ is empty, since this equation has only imaginary roots. To denote that a set $U$ is empty we write $U = \emptyset$. Here zero is crossed out by an oblique stroke in order to distinguish it from the number zero. In those cases, however, where there is no risk of confusion, we may write simply $U = 0$ without adding a stroke to the zero.

An empty set is usually considered as a subset of any set, i.e. if $U = \emptyset$, for any set $A$ the the relation $U \subseteq A$ is correct, and also $U \subset A$, when $A$ itself is not empty.

A linear point set is said to be *bounded* if it is a subset of some segment.

**2.** If in the series of natural numbers $1, 2, \ldots, n, \ldots$ we associate with each natural number some number $a_n$, according to a definite law, we obtain a numerical *sequence* whose *elements* (or *terms*) are numbered by all the natural numbers and are arranged in the order of the magnitude of these numbers. The sequence is represented by writing out its elements

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

or, briefly, by the symbol $\{a_n\}$. For example, the sequence $\left\{\frac{1}{n^2}\right\}$ consists of the elements $1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \ldots$, i.e. $a_1 = 1$, $a_2 = \frac{1}{4}$, $a_3 = \frac{1}{9}$, $a_4 = \frac{1}{16}$, etc.

More often the sequence is given by a formula which defines its *general term* $a_n$ as a function of the natural number $n$. Thus the formula $a_n = \frac{n-1}{n+1}$ defines the sequence

$$\{a_n\} = \left\{\frac{n-1}{n+1}\right\} = 0, \frac{1}{3}, \frac{2}{4}, \frac{3}{5}, \frac{4}{6}, \ldots;$$

the formula $a_n = 2^n$, the sequence

$$\{a^n\} \equiv \{2^n\} \equiv 2, 4, 8, 16, \ldots$$

etc.

Another common method of representing sequences is by *recurrence relations*, by means of which each successive term of the sequence is determined from the preceding. For example, $a_1 = 1$, $a_n = a_1 + a_2 + \cdots + a_{n-1}$, are recurrence relations which define the sequence:

$$a_2 = a_1 = 1,$$
$$a_3 = a_1 + a_2 = 2,$$
$$a_4 = a_1 + a_2 + a_3 = 4,$$
$$a_5 = a_1 + a_2 + a_3 + a_4 = 8,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

A sequence may also be given by any other method, provided it is always possible so to calculate any one of its terms. For instance the sequence $\{a_n\}$ is fully defined by the statement that $a_n$ is equal to the $n$th digit of the decimal expansion of the number $\pi = 3{\cdot}1415926536 \ldots$, from which it is obvious that $a_1 = 3$, $a_2 = 1$, $a_3 = 4$, $a_4 = 1$, $a_5 = 5$, $a_6 = 9$, ... To find each term of the sequence we need only calculate the number $\pi$ to a sufficiently high degree of accuracy.

It must be borne in mind that terms of a sequence may be equal to each other, while the elements of a set are always assumed to be different. Therefore it is necessary to distinguish a sequence from the set, which consists of the elements of the sequence. This difference is important even in the case where all the elements of the sequence are different: the set is defined by the elements of which it consists, but for the definition of the sequence it is necessary to know also in what order these elements are arranged. For instance the sequences $\{1, 2, 3, 4, 5, 6, \ldots\}$ and $\{2, 1, 4, 3, 6, 5, \ldots\}$ are different sequences but consist of exactly the same set of elements.

**3.** Suppose that $A$ and $B$ are two sets and that one and only one element of the set $B$ corresponds to each element of the set $A$. If each element of the set $B$ participates in this relationship, i.e. one and only one element of the set $A$ stands in this relation to

each element of the set $B$, we say that a *one-to-one correspondence* has been established between the sets $A$ and $B$.

For instance, to each element of a sequence all of whose elements are different, there corresponds one and only one integer—its number in the sequence; on the other hand, each integer is the sequence number of some particular term of the sequence, so that we have a one-to-one correspondence between the set of elements of the sequence and the set of natural numbers.

Sets, between which we can establish a one-to-one correspondence, are said to be *equivalent* or *of equal power* (having the same *power*). The equivalence of sets $M$ and $N$ is written in the form $M \sim N$.

It is clear that finite sets (i.e. those which consist of a finite number of elements) are of equal power if, and only if, they contain the same number of elements. No proper subset of a finite set can be equivalent to the whole set.

Conversely for infinite sets the following property holds: *for any infinite set we may find a proper subset, which is equivalent to the set itself*. This property may be used to define what we mean by an infinite set.

The simplest infinite sets are those which are equivalent to the set of natural numbers. Such sets are called *denumerable*. The power of a denumerable set is denoted by $\aleph_0$ (read as aleph-null or aleph-zero).

As an example of a denumerable set we may take a sequence, all of whose elements are different. On the other hand, the possibility of establishing a one-to-one correspondence between some set and the set of natural numbers shows that the elements of any denumerable set may be arranged in the form of a sequence.

The set of all integers (positive and negative, including zero) is denumerable. To be convinced of this it is sufficient to arrange them in the form of the sequence

$$0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots$$

The set of all rational numbers† of the segment [0, 1] is denumerable. In fact the rational numbers of the segment [0, 1] may be arranged in a sequence in the following way; we write them in

† The number $\alpha$ is rational if it may be put in the form $\alpha = p/q$, where $p$ and $q$ are integers.

the order of magnitude of the denominators and, with the same denominator, in the order of the numerator, omitting reducible fractions. We thus obtain the sequence

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \cdots$$

It is clear that each natural number $x$, satisfying the condition $0 \leqslant x \leqslant 1$, occurs in this sequence once and once only.

It is easy to prove that even the *set of all rational numbers is denumerable.* For the proof we use the following geometrical illustration. Let us represent a rational number by a point of a plane, setting off on the axis $Ox$ the numerator, and on the axis $Oy$ the denominator, considering that the latter is always positive (for whole numbers, naturally, the denominator must be considered as unity). The set of rational numbers is then represented by a lattice, whose points are distributed in the first and second quadrants of the plane $Oxy$ (Fig. 1). Now there is no difficulty in numbering them in succession, for instance, along a spiral as shown in Fig. 1 by arrows, omitting the reducible fractions, so that each rational number occurs in this sequence exactly once.

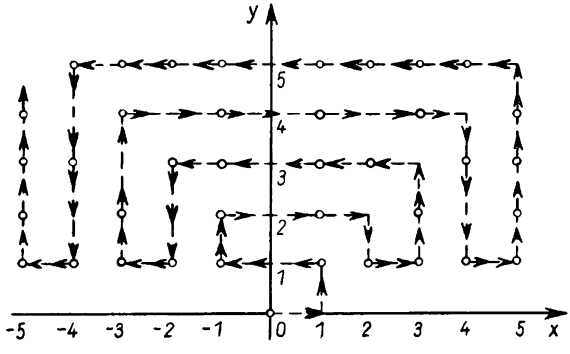


FIG. 1

The role of denumerable sets as the simplest of infinite sets is particularly stressed by the theorems:

*Any infinite set contains a denumerable subset.*

*Any subset of a denumerable set is either finite or denumerable.*

**4.** Let $A$ and $B$ be two arbitrary sets. The following relations between them are possible.

(1) The sets $A$ and $B$ are equivalent, i.e. they have the same power.

(2) One of the sets, set $A$ for example, is equivalent to some subset of $B$, but not equivalent to $B$ itself. In this case we say that *the power of the set $B$ is greater than the power of $A$*. Conversely, if the set $B$ is equivalent to a subset of the set $A$, but not equivalent to $A$, the power of $B$ is less than the power of $A$.

(3) The set $A$ is equivalent to some subset of $B$, and the set $B$ is equivalent to some subset of $A$. *Then the sets $A$ and $B$ are equivalent to each other* (the Cantor–Bernstein theorem).

The case where there is no subset in the set $A$, which is equivalent to $B$, and in the set $B$ no subset equivalent to $A$, is *impossible*. The proof of this, however, turns on the so-called *axiom of Zermelo*, whose application some mathematicians consider it necessary to avoid.

Infinite sets are not exhausted by those which are denumerable. A very important example of a set of greater power is the set of all the real numbers of a segment. *The set of all real numbers of the segment* [0, 1] *is not denumerable.*

The proof of this statement is carried out most easily by the method of contradiction. Each real number of the segment [0, 1] may be written in unique form as a decimal fraction $0 \cdot a_1 a_2 \ldots$, which is not a periodic fraction of period 9, where $a_n$ ($n = 1, 2, \ldots$) are the corresponding decimal digits.† Let us assume, in contradiction to the statement, that the set of real numbers of the segment [0, 1] is denumerable. In such a case we may arrange it as a sequence

$$0 \cdot a_{11}a_{12}a_{13} \ldots a_{1n} \ldots .$$
$$0 \cdot a_{21}a_{22}a_{23} \ldots a_{2n} \ldots .$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$0 \cdot a_{n1}a_{n2}a_{n3} \ldots a_{nn} \ldots .$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$

so that $a_{ik}$ denotes the $k$th decimal digit of the $i$th real number, according to the assumption that this sequence contains all the real numbers $x$, for which $0 \leqslant x \leqslant 1$.

† The only exception is the number 1, which must be put in the form $1 = 0 \cdot 999 \ldots$

We easily reach a contradiction, however, by constructing an infinite decimal fraction which is *not included* in the given sequence. For this it is sufficient to choose numbers $b_n$ quite arbitrarily, so long as $b_n \neq a_{nn}$ $(a = 1, 2, \ldots)$ and the number $0 \cdot b_1 b_2 b_3 \cdots b_n \ldots$, is not a periodic fraction with period 9. This number is not contained in our sequence. In fact, if it has the number $p$ in it then for all $k$ the equation $a_{pk} = b_k$ would hold, which with $k = p$ would be automatically untrue, since in our construction $b_p \neq a_{pp}$. The contradiction arrived at proves the non-denumerability of the set of real numbers.

We may always establish a one-to-one correspondence between two segments of any length by means of the projection shown in Fig. 2. Therefore the *set of all points of any segment is non-denumer-*

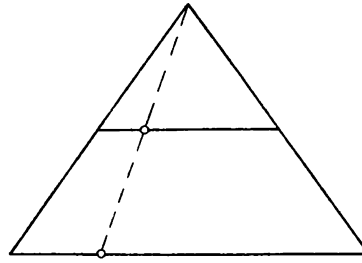

FIG. 2

*able.* Sets which have the same power as the set of all points of the segment are called sets of the *power of the continuum.* The power of the continuum is denoted by $\aleph$ (aleph) or $\mathfrak{C}$.

Sets which are all of increasing power may be obtained by the use of the theorem *the set of all subsets of a non-empty set has a power greater than the power of the original set.*

If a finite set consists of $N$ elements it has $2^N$ different subsets. Therefore, if the power of a set is denoted by $a$, the power of the set of all its subsets is represented by $2^a$.

*If a set is denumerable, the set of all its subsets has the power of the continuum*: $2^{\aleph_0} = \aleph$.

5. Suppose that $A$ and $B$ are two arbitrary point sets on a straight line. The *sum* (or *union*) of the sets $A$ and $B$ is the set $S$,

which consists of all the elements, belonging to at least one of the sets $A$ or $B$. The sum of the sets is represented by the symbol

$$S = A \bigcup B.$$

Also in those cases where there is no risk of confusing this operation with other operations, we often write $S = A + B$.

By the *difference* between the sets $A$ and $B$ we understand the set $E$, consisting of all those elements of the set $A$, which are not contained in $B$. Such a difference is denoted by $A \setminus B = E$. Here the set $B$ need not necessarily be a subset of $A$. If $B \subseteq A$, then we often write $E = A - B$. The set $E$ is called the *complement* of the set $B$ with respect to $A$ and if $A$ denotes a segment or the whole line, then simply the *complement*. The complement of the set $B$ is represented by $CB$.

The *intersection* (or *product*) of the sets $A$ and $B$,

$$T = A \bigcap B \quad \text{or} \quad T = A \,.\, B = AB,$$

is the set $T$, which consists of all the elements which belong both to the set $A$ and to the set $B$. If this set $T$ is empty, we say that the sets $A$ and $B$ *do not intersect*, or *are disjoint*.

For instance, if $E_1$ denotes the set of natural numbers and $E_2$ the set which consists of the negative integers and zero, the sum $E = E_1 + E_2$ coincides with the set of all integers. The difference between the interval $(0, 1)$ and the segment $[\frac{1}{3}, \frac{2}{3}]$ consists of two intervals $(0, \frac{1}{3})$ and $(\frac{2}{3}, 1)$, so that we may write

$$(0,\ 1) \setminus \left[\frac{1}{3},\ \frac{2}{3}\right] = \left(0,\ \frac{1}{3}\right) \cup \left(\frac{2}{3},\ 1\right)$$

or

$$(0, 1) - \left[\frac{1}{3}, \frac{2}{3}\right] = \left(0,\ \frac{1}{3}\right) + \left(\frac{2}{3}, 1\right).$$

If $A \supseteq B$ then $A \bigcup B = A$ and $A \bigcap B = B$.

Thus $(0, 1) \bigcap [\frac{1}{3}, \frac{2}{3}] = [\frac{1}{3}, \frac{2}{3}]$. The sets $E_1$ and $E_2$ defined above, have an empty intersection.

The sum and intersection of a system of sets $\{E_\alpha\}$ are denoted respectively by

$$\bigcup_\alpha E_\alpha \quad \text{or} \quad \sum_\alpha E_\alpha$$

and

$$\bigcap_\alpha E_\alpha \quad \text{or} \quad \prod_\alpha E_\alpha.$$

For instance, the set of all positive irrational numbers may be put in the form of a sum

$$E = \bigcup_{n=1}^{\infty} E_n = \sum_{n=1}^{\infty} E_n,$$

where $E_n$ denotes the set of all the irrational numbers of the segment $[n - 1, n]$.

The power of the sum of sets possesses the following properties:

(a) *the sum of a denumerable set of mutually disjoint finite sets is a denumerable set*;

(b) *the sum of a finite or denumerable set of denumerable sets is denumerable*;

(c) *if $E$ is an infinite set and $A$ a finite or denumerable set, then $E$ is equivalent to $E \cup A$. If the set $E$ is non-denumerable, then it is equivalent also to $E \setminus A$*;

(d) *if the elements of a set are determined by a finite number of indices, each of which may take a denumerable set of values, then such a set is denumerable*;

(e) *the sum of a finite or denumerable set of sets, which have the power of the continuum, has the power of the continuum.*

The above properties enable us to draw a series of useful conclusions. For instance:

The set of all points of a plane (or of three-dimensional space), all of whose coordinates are rational, is denumerable.

The set of all polynomials of any degree with rational coefficients is denumerable.

The set of all algebraic numbers, i.e. numbers which are roots of algebraic equations with rational coefficients, is denumerable.

The set of all real numbers has the power of the continuum.

The set of all transcendental numbers, i.e. real numbers which are not algebraic, has the power of the continuum.

It may appear that the set of points of a plane has a greater power than the set of points of a line. In fact this is not so; *the set of all points of a plane*† *has the power of the continuum.*

Passing from the consideration of sets to their complements, the operations of summation and intersection are interchanged in the sense that, for a system of sets $\{E_\alpha\}$, *the complement of the sum*

† And even the set of all points of Euclidean space of $n$ dimensions.

*of the sets is equal to the intersection of the complements of the selected terms*:

$$C(\sum_{a} E_a) = \prod_{a} CE_a,$$

and conversely, *the complement of the intersection of the sets is equal to the sum of the complements of the individual sets:*

$$C(\prod_{a} E_a) = \sum_{a} CE_a.$$

6. *A neighbourhood of a point* $a$ is any interval with its centre at the point $a$; *the* $\varepsilon$*-neighbourhood* is the interval $(a - \varepsilon, a + \varepsilon)$.

The point $a \in F$ is an *isolated point* of the set $E$ if there exists an $\varepsilon$-neighbourhood of $a$ which does not contain points of $E$, different from the point $a$.

The point $a$ is called a *limit point* of the set $E$, if any neighbourhood of the point $a$ contains at least one point of the set $E$, different from $a$. The point $a$ may or may not belong to $E$.

It immediately follows from the definition that *any neighbourhood of a limit point contains an infinite set of points of* $E$.

If a linear set $E$ is not bounded, then it has points outside any segment. It is possible that the set $E$ may have points to the right of any point of the real axis or to the left of any of its points. In the first case we say that the set $E$ is *unbounded on the right* and that it has $+\infty$ for its limit point; in the second case that it is *unbounded on the left* and that $-\infty$ is the limit point of $E$. In those cases where the set $E$ has points both to the right and left of any point of the axis, we say that the set has as limit points both $+\infty$ and $-\infty$.

The point $a$ is called *the limit point of the sequence* $\{a_n\}$ if any neighbourhood of $a$ contains elements of the sequence for a sufficiently large value of $n$. If the sequence is unbounded to the right or left it is said that its limit point is $+\infty$ or $-\infty$ respectively.

As we have already shown, terms of the sequence may be repeated. We must, therefore, distinguish between the limit points of a sequence and the limit points of the set of its elements. The limit point of a set of the elements of a sequence is certainly a limit point of the sequence. The opposite conclusion is not true. For instance the sequence 0, 1, 0, 1, ... has two limit points 0 and 1, but the set of the elements of this sequence is finite and can have no limit points.

If the sequence has only one limit point then this point is *the limit of the sequence*, i.e. the sequence *converges to this point*. In order that the point $a$ may be a limit point for the set $E$, it is necessary and sufficient that there should exist a sequence $\{x_n\} \subset E$ such that $x_n \neq x_k$ with $n \neq k$ and $\lim_{n\to\infty} x_n = a$.

The set of all limit points of the set $E$ is called the *derived set* and is denoted by $E'$. Adding the derived set $E'$ to the set $E$, we obtain the *closure* of the set $E$, which is denoted by $\overline{E}$;

$$\overline{E} = E + E'.$$

The set $E$ is said to be *closed* if $E' \subseteq E$. Alternatively, *a closed set contains all its limit points and coincides with its closure*: $\overline{E} = E$. Each of these properties may be used as a definition of a closed set.

We shall say that the point $a$ is *the least upper bound point* (or simply the *least upper bound*) of a set $E$, if for any point $x \in E$ the inequality $x \leqslant a$ is true, and for any $\varepsilon > 0$ a point $y \in E$ can be found for which $y > a - \varepsilon$. If $a$ is the least upper bound of $E$ we say that $a$ is the *supremum* of $E$ and we write $a = \sup E$.

With $a \in E$ we may for any $\varepsilon$ assume that $y = a$, so that the upper bound may be an isolated point; if $a \overline{\in} E$ the point $a$ is necessarily a limit point. For example, the point $a = 1$ is the upper bound of the segment $[0, 1]$ and the interval $(0, 1)$ and the point $a = 2$ of the set $\left\{2 - \frac{1}{n}\right\}$ $(n = 1, 2, \ldots)$. If the set includes the extreme right-hand point the upper bound of the set coincides with this point. On the other hand if $a = \sup E$ belongs to the set $E$, the set $E$ includes the extreme right-hand point.

Similarly the greatest *lower bound point* (greatest *lower bound*) of the set $E$ is the point $a$ such that for any $x \in E$ the inequality $x \geqslant a$ is true, and for any $\varepsilon > 0$ there can be found $y \in E$, for which $y < a + \varepsilon$. The lower bound of the set $E$ is also called the *infimum* of $E$ and is denoted by $\inf E = a$.

*A bounded linear point set always has a least upper and a greatest lower bound.*

Let the set $E$ be closed and bounded. Then $\inf E \in E$ and $\sup E \in E$, i.e. a bounded closed set always contains its extreme right and left points.

We shall denote a bounded linear point set by $E$. The least upper bound of the derived set $E'$ is called *the upper limit of the*

*set L*, if it is not empty. Since *the derived set is always closed* the upper limit of $E$ belongs to $E'$. *Thus the upper limit of a bounded set $E$ is the extreme right limit point of $E$.* The upper limit of $E$ is represented by $\limsup E$ or $\overline{\lim} E$.

Similarly the *extreme left limit point*, i.e. the lower bound of the derived set, is called the *lower limit of* a bounded set, if it is not empty. The lower limit of the set $E$ is denoted by $\liminf E$ or $\underline{\lim} E$.

If the set $E$ is not bounded on the right we say that its upper limit is $+\infty$, $\limsup E = +\infty$. The lower limit of the set may similarly be $-\infty$, $\liminf E = -\infty$, if it is not bounded on the left; it may be any finite point if the set $E$ is bounded on the left and has limit points different from $+\infty$; finally it may be $+\infty$, $\liminf E = +\infty$, if the set has no other limit points besides $+\infty$. Actually also if the set is not bounded on the left, but is bounded on the right and has no limit points besides $-\infty$, then $\limsup E = \liminf E = -\infty$.

The upper and lower limits of a sequence are defined in a similar way. The *upper* (*lower*) limit of a sequence $\{a_n\}$ is its extreme right (left) limit point or $+\infty$ ($-\infty$), if it is not bounded to the right (left). The sequence which has $+\infty$ as its upper limit may have $-\infty$, an arbitrary point of the axis, or $+\infty$ as its lower limit.

The upper limit of a sequence is denoted by the symbol

$$\limsup_{n\to\infty} a_n \quad \text{or} \quad \overline{\lim_{n\to\infty}} a_n,$$

and the lower by the symbol

$$\liminf_{n\to\infty} a_n \quad \text{or} \quad \underline{\lim_{n\to\infty}} a_n.$$

For the sequence $\left\{\frac{1}{n}, \frac{n-1}{n}\right\}$ i.e. $1, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{4}{5}, \cdots$,

$$\limsup_{n\to\infty} a_n = 1, \quad \liminf_{n\to\infty} a_n = 0.$$

If the upper and lower limits of a sequence coincide, the sequence has a unique limit point to which it converges.

The upper limit of a set must be distinguished from its upper bound. While all elements of the set are situated to the left of its upper bound, there may exist an infinite subset of them, situated to the right of the upper limit.

*Any finite bounded set has at least one limit point* (the Bolzano–Weierstrass theorem).

The point $a$ is called an *inner point* of the set $E$ if there exists a neighbourhood of $a$, all of whose points belong to $E$. The set, all of whose points are inner points, is called *an open set* or *domain* (to stress that the term applies to a linear set we also say *"linear domain"*). An example of an open set is the arbitrary interval $(a, b)$.

*The complement of a closed set is open.* Conversely *the complement of an open set is closed.* Let $F$ be a closed set and $a \in CF$. In such a case there exists an $\varepsilon$-neighbourhood of the point $a$, free from points of $F$, since, if in any neighbourhood of $a$ there were points of $F$, $a$ would be the limit point of $F$ and because of the closure of $F$ it would belong to $F$ and not $CF$. Consequently this $\varepsilon$-neighbourhood must belong entirely to $CF$. Thus any point of $CF$ is an inner point and hence $CF$ is an open set. Now let $G$ be open and $a$ be a limit point of the set $CG$. Then $a$ cannot belong to $G$, otherwise there could be found a neighbourhood of $a$, wholly included in $G$ and not containing any points of $G$. Hence $a$ would not be a limit point for $CG$. Consequently all limit points of $CG$ belong to $CG$ so that this set is closed.

Because of the connection established between open and closed sets, it is sufficient to consider the basic properties and structure of only one of these classes.

Operations on closed sets possess the following properties:

(a) *The intersection of any set of closed sets is a closed set.*

(b) *The sum of a finite set of closed sets is a closed set.*

The sum of a denumerable set of closed sets may not even turn out to be closed. We may easily be convinced of this by noting, for instance, that a set which consists of one point is closed, while the set of the elements of the sequence $\left\{\frac{1}{n}\right\}$, which is the sum of a denumerable set of closed sets, does not contain the limit point 0, and is therefore not closed.

Turning to complements, we see that

(c) *The sum of any set of open sets is an open set.*

(d) *The intersection of a finite set of open sets is an open set.*

Hence, in particular, it follows that the sum of a finite or denumerable set of mutually disjoint intervals is an open set. Also

the converse statement is true, which exhaustively explains the structure of open sets on a straight line:

*Any bounded open linear set $G$ may be expressed in the form of a sum of a finite or denumerable set of mutually disjoint intervals*

$$G = \sum_k (a_k, b_k).$$

If we turn to complements, we may state that *any bounded closed set $F$ is either a segment or is obtained from some segment by the subtraction of a finite or denumerable set of mutually disjoint intervals*

$$F = [a, b] - \sum_k (a_k, b_k).$$

These intervals bear the name *complementary or adjoint intervals of the closed set $F$.*

**7.** The set $E$ is said to be *perfect* if it coincides with its derived set, $E = E'$. It is obvious that a perfect set is closed and has no isolated points. Conversely, *a closed set devoid of isolated points, is perfect.*

The point $a$ is called a *condensation point* of the set $E$ if any neighbourhood of the point $a$ contains a *non-denumerable* set of points of $E$. A condensation point is always a limit point but a limit point may not be a condensation point. *Any non-denumerable set has at least one condensation point belonging to it.* It is helpful to turn our attention to the difference between this statement and the Bolzano-Weierstrass theorem.

*A non-empty perfect set has the power of the continuum; the power of a non-empty closed set may be either the continuum, or denumerable, or finite.*†

*A non-denumerable closed set may always be given in the form of the sum of a perfect set and is not greater than a denumerable set of points* (the Cantor-Bendixson theorem).

Adjoint intervals of a perfect set have no common ends.

† By the same token we may state that a closed set cannot have a power intermediate between denumerable and that of the continuum. At the present time mathematicians assume that sets of intermediate power cannot in general exist, but this has not up to now been proved. Such a statement in the form given, or in an equivalent form, is called the *continuum hypothesis.*

8. A set $E$ is said to be *dense in itself* if $E \subseteq E'$, *dense on the set* $R$ if $E' \supseteq R$, and *everywhere dense* if $R$ is a segment which contains $E$ or the whole line. The set $E$ is said to be *nowhere dense* if it is not dense on any segment. Alternatively, a set $E$ is nowhere dense if, on each segment $[a, b]$, we can find a segment $[a', b'] \subset [a, b]$, which does not contain one point of $E$.

The complement of a nowhere dense set is necessarily everywhere dense. The complement of an everywhere dense set is, however, not only not bound to be nowhere dense, but itself may even turn out to be everywhere dense. Thus the set of all rational points of the segment [0, 1] is everywhere dense at the same time as its complement.

Basic interest is attached to the example of a nowhere dense perfect set first constructed by G. Cantor. This set is built up in the following way. We divide the segment $U = [0, 1]$ into three equal parts by the points $\frac{1}{3}$ and $\frac{2}{3}$ and remove from $U$ the middle interval $(\frac{1}{3}, \frac{2}{3})$. We again divide each of the remaining segments $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ into three equal parts and remove their middle intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$.

FIG. 3

We treat each of the four remaining segments in exactly the same way; we divide them into three equal parts and remove the middle interval (Fig. 3). Continuing the process indefinitely, we note that from the segment $U$ we can count a denumerable set of mutually disjoint intervals, which have no common ends, either with each other or with the basic segment [0, 1]. Hence it follows that the remaining set $P = U - \sum_n (a_n, b_n)$ is perfect. This set is called a *Cantor perfect set.*

The process described is typical of the construction of nowhere dense perfect bounded sets. Every perfect, nowhere dense set may be constructed by means of a similar process, so that it is *similar* to a Cantor perfect set. The latter means that we may establish a one-to-one correspondence between two nowhere dense perfect sets which retains the order of succession of the points.

If the given set is obtained in the form of the sum of a finite or denumerable set of nowhere dense sets, it is called a set of the

*first category*. Sets which are not of the first category are called sets of the *second category*. The sum of a denumerable set of sets of the first category is itself a set of the first category. Since a segment is not a set of the first category, the complement of a set of the first category cannot itself be a set of the first category. Such sets are called sets *strictly of the second category* or *of the second category in the restricted sense.*

## § 2. THE LEBESGUE MEASURE OF LINEAR SETS

**1.** The idea of power which enables us to some extent to distinguish infinite sets as "greater" or "less" is nevertheless insufficient for the study of linear sets. In fact, segments of any length are equivalent to each other and have the power of the continuum. Naturally there arises a necessity to establish characteristics of a set, which reflect its extent and generalize the idea of length. The *measure* $(m)$ of a linear set is such a characteristic.

For open and closed sets the following definitions are sufficient:

(a) The measure of an interval is its length, $m(a, b) = b - a$.

(b) The measure of a bounded open set $G$ is the sum of the lengths of its constituent intervals

$$mG = \sum_k (b_k - a_k),$$

where $(a_k, b_k)$ $(k = 1, 2, \ldots)$ are intervals which make up $G$, while the set may be both finite and denumerable. In the first case the sign $\sum_k$ must stand for $\sum_{k=1}^{n}$, and in the second case $\sum_{k=1}^{\infty}$.

(c) The measure of a bounded closed set $F$ is the difference

$$mF = b - a - mCF,$$

where $[a, b]$ is some segment which contains the set $F$, and $CF$ the complement of the set $F$ with respect to this segment. It is not difficult to see that the value of $mF$ does not depend on the choice of the segment $[a, b] \supseteq F$.

Let us consider as an example the measure of the Cantor perfect set, constructed in § 1, para. 8. The sum of the lengths of adjoint intervals is

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots + 2^n \cdot \frac{1}{3^{n+1}} + \cdots = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = 1,$$

so that the *measure of a Cantor perfect set is zero.*

These definitions put in correspondence with each open or closed set $E$, some definite number—*the measure of* $E$, denoted by $mE$ (sometimes the notations mes $E$ or $|E|$ are used) and possessing obvious properties similar to those of length. There are not sufficient of them, however, to put a numerical value of the measure in correspondence with a set of a rather complex nature.

For the widest class of set we may obtain the measure with the help of a construction put forward by H. Lebesgue. This measure is called the *Lebesgue linear measure.*

**2.** Let $E$ be an arbitrary, totally disconnected set, lying inside the segment $[a, b]$. If $G$ is an open set, belonging to the same segment and containing $E$, $G \supset E$, it is said that the set $E$ is *covered* by the open set and $G$ is called a *covering* of $E$.

We shall denote by $mG$ the measure of the set $G$, i.e. the sum of the lengths of its constituent intervals. *The outer measure of the set $E$ is the lower bound of the measures of the open sets covering $E$.* Representing the other measure of $E$ by $m_eE$ we have by definition,

$$m_eE = \inf_{G \supset E} mG.$$

The set $E$ is said to be *measurable in the Lebesgue sense* if

$$m_eE + m_eCE = b - a.$$

In this case for the measure of $E$ we use its outer measure and write $mE = m_eE$. The difference,

$$b - a - m_eCE = m_iE,$$

is called the *inner measure*† *of the set $E$. Therefore the set $E$ is measurable when, and only when, $m_eE = m_iE$. The Lebesgue measure of a measurable set is the common value of its outer and inner measures.*

The properties of measurable sets are expressed by the following theorems:

(a) *A bounded denumerable set is measurable and its measure is zero.*

† The inner measure of the set $E$ may also be defined as *the upper bound of the measures of the closed sets contained in $E$,* $m_iE = \sup_{F \subset E} mE$.

(b) *A bounded set $E$, which is the sum of a finite or denumerable set of measurable sets, $E = \sum_k E_k$, is measurable. If the terms are mutually disjoint then*

$$mE = \sum_k mE_k.$$

(c) *The difference between two measurable sets $E = E \setminus E_2$, is measurable.* If $E_1 \supseteq E_2$ then

$$mE = mE_1 - mE_2.$$

(d) *The intersection of a finite or denumerable set of measurable sets is measurable.*

(e) Let $E_1, E_2, E_3, \ldots$ *be measurable and* $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$ *If the sum* $\sum_{n=1}^{\infty} E_k = E$ *is bounded, then*

$$mE = \lim_{n\to\infty} mE_n.$$

(f) *Let $E_1, E_2, E_3, \ldots$ be measurable, $E_1 \supseteq E_2 \supseteq E_3 \supseteq \ldots$* and $E = \prod_{n=1}^{\infty} E_n$. Then

$$mE = \lim_{n\to\infty} mE_n.$$

*All bounded open and closed sets are measurable in the Lebesgue sense.* The value of the Lebesgue measure for these sets coincides with the value of the measure defined in § 2, para 1. *Moreover, any bounded set, which may be obtained from open or closed sets by the use of finite or denumerable sets of operations of addition and intersection, is measurable.*

Any set thus obtained is called a *Borel* set ($B$-set) *or a set measurable B.* Thus any set, measurable $B$, is measurable in the Lebesgue sense. The converse theorem is not true.

**3.** A measurable set $E$ of positive measure is said to be *reduced* if any interval containing at least one point of $E$ also contains a part of $E$ of positive measure. Any measurable set of positive measure contains a reduced set of the same measure. We may construct one by forming the difference between the set $E$ and all those intervals with rational ends, which contain parts of the set $E$ of measure zero.

The structure of a measurable set of positive measure is described by the theorem:

*Any measurable set of positive measure may be put in the form of a sum of a denumerable set of mutually disjoint, reduced, perfect sets and a residual set of measure zero.*

On considering a similar assertion on addition we can obtain an approximation to a measurable set, not only from within but also from without:

*For any measurable set $E$ there exist B-sets $E_1$ and $E_2$ of the same measure, one of which is contained in the set $E$ and the other of which contains it,* i.e.

$$E_1 \subseteq E \subseteq E_2 \quad \text{and} \quad mE_1 = mE_2 = mE.$$

Alternatively, except for a set of measure zero any measurable set may be considered as a $B$-set. If we limit ourselves to consideration within sets of as small measure as we please, a measurable set may be considered as perfect, including simply the sum of a finite number of segments. To put it more accurately:

*For any measurable set of positive measure, and any number $\varepsilon > 0$, there exists a finite number of segments such that the points of the set not occurring in them, and the points of its complement occurring in them, together form a set whose measure is less than $\varepsilon$.*

For the study of the local structure of measurable sets of positive measure we use essentially the idea of density points.

Let $E$ be a measurable set, lying on the segment $[a, b]$ and $x_0 \in [a, b]$ an arbitrary point, belonging to or not belonging to $E$. Let us denote by $\Delta$ the interval of length $\delta$, which contains the point $x_0$, and by $m(\Delta E)$ the measure of the part of $E$ belonging to this interval. The point $x_0$ is called *a density point* of the set $E$ if

$$\lim_{\delta \to 0} \frac{m(\Delta E)}{\delta} = 1.$$

If the above limit is zero the point $x_0$ is said to be a *point of rarefication.*

For a set, whose measure is equal to the length of the segment which contains it (such sets are called sets of *complete* measure), each point of the segment is a density point. Conversely, if the measure of a set is zero, each point of the segment will be a point of rarefication for this set. *For a measurable set $E \subset [a, b]$, whose measure satisfies the conditions $0 < mE < b - a$, all the points, except for a set of measure zero, are density points of $E$, and all points*

*of its complement, apart from a set of measure zero, are points of rarefication of E.*

If all points of some set, apart from points which form a set of measure zero, possess some property, we say that this property belongs to *almost all* points, or that this property holds *almost everywhere.* Using this terminology, we may say that *almost all points of a measurable set are its density points.* This fact shows that a set of intermediate measure is arranged along the segment, not uniformly, but in a non-uniform manner, in groups, densely in some places and sparsely in others, similar to the system of segments.

The resemblance between a measurable set and a segment is substantiated by the following statements:

*If measurable sets $E_1$ and $E_2$ have $x_0$ as a density point, then they intersect and $x_0$ is a density point of the intersection.*

*If the sets $E_1$ and $E_2$, which lie on the segment $[a, b]$, are measurable and the sum $S$ of their measures exceeds $b - a$, then they intersect and the measure of their intersection is greater than $S - (b - a)$.*

**4.** Let $E_1, E_2, \ldots, E_n, \ldots$ be a sequence of linear sets. *The upper limit* of a given sequence of sets is the set of all points, such that each of them enters into an infinite number of sets of the sequence. The set of all points, each of which is contained in all the sets of the sequence, except for a finite number, is said to be the *lower limit* of the sequence of sets. The upper and lower limits of the sequence of sets are represented respectively by $\overline{\lim\limits_{n\to\infty}} E_n$ and $\underline{\lim\limits_{n\to\infty}} E_n$ or $\limsup\limits_{n\to\infty} E_n$ and $\liminf\limits_{n\to\infty} E_n$. Alternatively, the set, which is the lower limit of a sequence, consists of points belonging to all the sets of the sequence, commencing with a certain number, and the set which is the upper limit consists of points, which belong to the sets with numbers as large as we please.

The upper and lower limits of a sequence of sets may be obtained by means of the operations of addition and intersection by the formulae

$$\underline{\lim_{n\to\infty}} E_n = (E_1 \,.\, E_2 \,.\, E_3 \,.\, \cdots) + (E_2 \,.\, E_3 \,.\, \cdots) + (E_3 \,.\, \cdots) + \cdots = \prod_{k=1}^{\infty} \sum_{n=k}^{\infty} E_n,$$

$$\overline{\lim_{n\to\infty}} E_n = (E_1 + E_2 + E_3 + \cdots) \,.\, (E_2 + E_3 + \cdots) \times (E_3 + \cdots) = \prod_{k=1}^{\infty} \sum_{n=k}^{\infty} E_n,$$

from which it follows, in particular, that if all sets of the sequence are measurable, then both the upper and lower limits are also measurable.

The upper and lower limits of a sequence of sets always exist. If they coincide, the sequence of sets is said to be *convergent*, and the set obtained is the *limit of the sequence of sets*:

$$\lim_{n\to\infty} E_n = \overline{\lim_{n\to\infty}} E_n = \underline{\lim_{n\to\infty}} E_n.$$

In the case where the upper and lower limits of the sequence differ, the sequence of sets is said to be *divergent*.

If a sequence of measurable sets $E_1, E_2, E_3, \ldots$, lying on a finite segment, converges to a set $E$, we write $E = \lim_{n\to\infty} E_n$.

*Monotonic* sequences of sets, i.e. sequences of sets enclosed one in another, are convergent. For a *monotonic increasing* sequence defined by the condition

$$E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq \cdots$$

the limit of the sequence is the sum

$$\lim_{n\to\infty} E_n = E_1 + E_2 + \cdots + E_n + \cdots = \sum_{n=1}^{\infty} E_n,$$

for a *monotonic decreasing* sequence for which by definition

$$E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq \cdots$$

the limit will be the intersection

$$\lim_{n\to\infty} E_n = E_1 \cdot E_2 \cdots E_n \cdots = \prod_{n=1}^{\infty} E_n.$$

It may happen that the limit of a monotonic increasing sequence of sets is the whole straight line, and the limit of a monotonic decreasing sequence is an empty set.

*With the convergence of the measure of $E_n$ to zero, with the unlimited growth of $n$, the measure of the lower limit of a divergent sequence of sets is equal to zero,* $m \underline{\lim}_{n\to\infty} E_n = 0$, but the upper limit, $\overline{\lim}_{n\to\infty} E_n$ may turn out to be of positive measure. If we assume in addition that the series from the measures of the sets of sequences converges, $\sum_{n=1}^{\infty} mE_n < \infty$ then $m \overline{\lim}_{n\to\infty} E_n = 0$.

**5.** Let $E \subset [a, b]$ be a measurable set and $\sum$ some system of intervals or segments, each point of $E$ belonging to at least one interval or segment of $\Sigma$. Then we say that *the set $E$ is covered by the system $\Sigma$*. We often use the following theorems on coverings.

(a) THE BOREL–LEBESGUE THEOREM. *If a bounded closed set $F$ is covered by an infinite system of intervals $\Sigma$, then from the system $\Sigma$ we may choose a finite number of intervals covering $F$.*

(b) *From the covering of a bounded closed set $F$ by a system of intervals, we may find a finite number of mutually disjoint intervals, which cover a part $E$ of the set $F$ with measure $mE \geqslant \frac{1}{2} mF$.*

(c) *If a bounded closed set $F$ is covered by a system of segments $\Sigma$, then from the covering we may choose a finite system of mutually disjoint segments, which cover a part $E$ of the set $F$ of measure $mE \geqslant \frac{1}{2} mF$.*

(d) VITALI'S THEOREM. *Let a bounded measurable set be covered by a system of segments $\Sigma$ so that each point of $E$ is covered by segments of length as small as we please. Then from the system $\Sigma$ we may choose a finite or denumerable set of disjoint segments $\{\Delta_k\}$, covering the set $E$ except for a set of measure zero*, i.e.

$$m\left[E \setminus \sum_k \Delta_k\right] = 0.$$

It is often convenient to use Vitali's theorem in the following form:

(e) *With the same assumptions as before, for any $\varepsilon > 0$ there exists a finite number of segments $\Delta_k$, which are mutually disjoint and such that*

$$m\left[E \setminus \sum_k \Delta_k\right] < \varepsilon.$$

In both forms of Vitali's theorem we may omit the assumption of the measurability of the set $E$, replacing the measure of the uncovered part of the set by its outer measure in the statement (e).

## § 3. BASIC CLASSES OF FUNCTIONS

**1.** Let us consider the arbitrary set $E \subset [a, b]$. Suppose that a definite value $y = f(x)$† corresponds to each value of $x \in E$. In such a case we say that a *function* is defined on the set $E$, and the set $E$ is called its *domain of definition*. The set $M$ of all possible values

† The value of $y$ may be both finite $-\infty < y < \infty$, and also infinite, $y = +\infty$ or $y = -\infty$.

of $y$ is called the *range of values* of the function or the *image of the set* $E$, which is written as the equation $M = f(E)$. If $y \in f(E)$, then any point $x \in E$, for which $f(x) = y$ is called the *inverse image* of $y$. The set of all such points is called the *complete inverse image* of $y$ and is represented by $f^{-1}(y)$.

Let the function $f(x)$ be defined on the set $E$.

The function $f(x)$ is said to be *bounded* on $E$ if the set $f(E)$ is bounded. It is said to be *finite* on the set $E$ if it takes, at each point $x \in E$, a finite value of $y = f(x)$, $-\infty < y < +\infty$. If the function $f(x)$ takes on a finite value almost everywhere on the set $E$, it is called *finite almost everywhere.*

We say that $f(x)$ *is continuous at the point* $x_0 \in E$ with *respect to the set* $E$ if, for any sequence $\{x_n\}$ of points, which belong to $E$ and converge to $x_0$, the sequence $\{f(x_n)\}$ converges to $f(x_0)$, i.e. from $\lim_{n\to\infty} x_n = x_0$, $x_n \in E$, it follows that

$$\lim_{n\to\infty} f(x_n) = f(x_0).$$

Alternatively, the function $f(x)$, defined on the set $E$, is continuous at the point $x_0 \in E$ with respect to $E$, if for any $\varepsilon > 0$ we can find $\delta > 0$ such that for all points $x \in E$, for which $|x - x_0| < \delta$, the relation

$$|f(x) - f(x_0)| < \varepsilon$$

is true.

The function $f(x)$, defined on a set $E$, is *uniformly continuous* on it if, for any $\varepsilon > 0$, we can find $\delta > 0$ such that, for any pair of points $x_1, x_2 \in E$, from $|x_2 - x_1| < \delta$ it follows that $|f(x_2) - f(x_1)| < \varepsilon$.

Functions which are continuous on closed sets have the following properties:

(a) *If a function* $f(x)$ *is continuous on a closed set* $F$, *then for any number* $c$ *the set of points* $F$, *for which* $f(x) \geqslant c$ *or* $f(x) \leqslant c$, *is closed.*

(b) *A function, which is continuous on a bounded closed set, is uniformly continuous on it.*

(c) *If a function* $f(x)$ *is continuous on a segment and takes values* $c_1$ *and* $c_2$ *on it, then it also takes on this segment any value* $c$ *which lies between them.*

(d) WEIERSTRASS' THEOREM. *If a function $f(x)$ is continuous on a bounded closed set, it is bounded on it and among its values there exist a least and a greatest.*

(e) *The continuous image of a closed set is a closed set. In particular, the continuous image of a segment is a segment.*†

Points of the set $E$, at which a function $f(x)$ defined on $E$ is not continuous with respect to $E$, are called *points of discontinuity* of this function.

Let a function $f(x)$ be defined on a set $E$ and let $x_0 \in E$. If for any sequence $\{x_n\}$, converging to $x_0$, for which $x_n \in E$ and $x_n > x_0$, the sequence $\{f(x_n)\}$ converges, we say that the function $f(x)$ has a *right hand limit* (or a *limit on the right*) at the point $x_0$, which is represented thus:

$$\lim_{\substack{x \to x_0 + 0 \\ x \in E}} f(x) = f(x_0 + 0).$$

Similarly, we define a *left-hand limit* (*limit on the left*). The function $f(x)$ is said to be continuous on the right (left), if

$$f(x_0) = f(x_0 + 0) \quad (f(x_0) = f(x_0 - 0)).$$

A point $x_0$ at which the limits on the left and right exist, but $f(x_0 - 0) \neq f(x_0 + 0)$, is called a *point of discontinuity of the first kind* or a *finite jump*. The difference $f(x_0 + 0) - f(x_0 - 0)$ called the *magnitude of the jump* at the point $x_0$. The differences $f(x_0 + 0) - f(x_0)$ and $f(x_0) - f(x_0 - 0)$ are called the *jump on the right* and *left* respectively. Points of discontinuity, distinct from finite jumps, are called points of discontinuity of the *second kind*. Points of discontinuity of a function may form a most extensive set. Thus the well-known *Dirichlet function*, defined on the segment [0, 1] by the condition

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is irrational,} \\ 1, & \text{if } x \text{ is rational,} \end{cases}$$

or, what is the same, by the formula

$$f(x) = \lim_{m \to \infty} \lim_{n \to \infty} (\cos 2\pi n! \, x)^m,$$

† We must bear in mind that the discussion relates to the mapping of linear sets onto linear sets.

is discontinuous at each point of the segment. The function

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is irrational,} \\ \dfrac{1}{q}, & \text{if } x = \dfrac{p}{q}, \quad \text{where } \dfrac{p}{q} \text{ is an irreducible rational fraction,} \end{cases}$$

is discontinuous at each rational point of the segment and continuous at irrational points.

2. The function $f(x)$, defined on a set $E$, is said to be *increasing* (or *non-decreasing*) on that set if, for any pair of points $x_1, x_2 \in E$, it follows from $x_2 > x_1$ that $f(x_2) \geqslant f(x_1)$. In the case where the latter relation has the form $f(x_2) > f(x_1)$ the function is *strictly increasing* (sometimes simply *increasing*).

Similarly, if for any $x_1, x_2 \in E$ it follows from $x_2 > x_1$ that $f(x_2) \leqslant f(x_1)$, the function is said to be *decreasing* or *non-increasing*, and in the case where $f(x_2) < f(x_1)$ *strictly decreasing*.

Increasing and decreasing functions are called *monotonic functions*, and those which are strictly increasing and strictly decreasing are *strictly monotonic*.

If a function $f(x)$ is increasing, the function $-f(x)$ will be decreasing and vice versa.

*Monotonic functions can have discontinuities only of the first kind.* Moreover, if $E$ is closed, *the set of points of discontinuity of the monotonic function are at most denumerable, and for any $\sigma > 0$ there exists only a finite number of points at which the increasing function has a jump greater than $\sigma$.*

Let $f(x)$ increase on the segment $[a, b]$. We shall define as a *jump function*, a function $f(x)$ on the assumption that for $x \in [a, b]$

$$s(a) = 0,$$

$$s(x) = [f(a + 0) - f(a)] + \sum_{x_k < x} [f(x_k + 0) - f(x_k - 0)] + [f(x) - f(x - 0)] \quad (a < x \leqslant b);$$

here $x_k$ are the abscissae of the jumps of $f(x)$.

Then the *difference between the increasing function and its jump function increases and is continuous, i.e. any increasing function may be put in the form of the sum of a continuous increasing function and a jump function*, which is obviously also non-decreasing. A similar theorem holds for a decreasing function.

**3.** Let the function $f(x)$ be defined and finite on a segment $[a, b]$. We divide the segment into parts by points

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

and construct for the given partitioning the sum

$$\mathbf{V} = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)| .$$

If the upper bound of the set of such sums is finite for all possible divisions it is called the *total variation* of the function $f(x)$ on the segment $[a, b]$ and is denoted thus:

$$\bigvee_a^b (f) = \sup \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)| ,$$

and the function $f(x)$ is called a *function of bounded variation on the segment.*

The simplest examples of functions of bounded variation are monotonic functions. *The total variation of an increasing function is equal to its increment on a given segment.* In fact since all terms in the sum which determines the variation are always positive for an increasing function, with any partitioning

$$\bigvee_a^b = f(b) - f(a).$$

It is clear that functions of bounded variation may be discontinuous. On the other hand, continuous functions do not necessarily

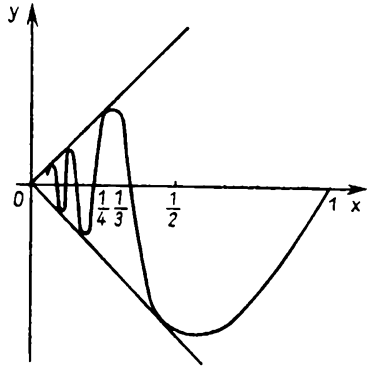


FIG. 4

have bounded variation. Thus the function $y = x \cos\left[\frac{\pi}{2x}\right]$, which is predetermined by the condition $y(0) = 0$, is continuous on the segment $[0, 1]$, but is not of bounded variation on that segment. Actually on dividing the segment $[0, 1]$, with fixed $n$ by the points

$$0 < \frac{1}{2n} < \frac{1}{2n-1} < \cdots < \frac{1}{3} < \frac{1}{2} < 1,$$

we find that for such a division the sum (Fig. 4)

$$\mathbf{V} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},$$

so that, with sufficiently large $n$ it may be as large as we please.

A very important class of function with bounded variation is formed by functions which satisfy a *Lipschitz condition.* A finite function $f(x)$ satisfies a Lipschitz condition on the segment $[a, b]$ if there exists a constant $K$ such that for any two points $x, y \in [a, b]$

$$|f(x) - f(y)| \leqslant K\,|x - y|.$$

In particular the function $f(x)$ satisfies a Lipschitz condition if at each point of $[a, b]$ it has a derivative which is bounded on the segment. In this case for $K$ we may use the upper bound of the values of $|f'(x)|$.

*A function $f(x)$, which satisfies a Lipschitz condition, on the segment $[a, b]$, has bounded variation on this segment and* $\overset{b}{\underset{a}{\mathbf{V}}}(f) \leqslant K(b-a)$.

Functions of bounded variation on a segment possess the following properties:

(a) *The sum, difference and product of two functions of bounded variation are functions of bounded variation. The quotient of two functions of bounded variation is also a function of bounded variation if the denominator $f(x)$ at each point satisfies the condition $f(x) \geqslant \geqslant \sigma > 0$.*

(b) *A function of bounded variation on a segment is bounded on it.*

(c) *If the segment $[a, b]$ is divided into a finite number of segments, the variation of the function on the segment is equal to the sum of the variations on all its parts. Hence it follows that if the segment $[a, b]$ can be divided into a finite number of parts, on each of which $f(x)$ is monotonic, $f(x)$ has bounded variation on $[a, b]$.*

(d) *In order that the function may have bounded variation on the segment* $[a, b]$, *it is necessary and sufficient that it should be possible to give it in the form of the difference between two functions which are increasing on* $[a, b]$. Using the properties of monotonic functions, from this we immediately conclude that:

(e) *A function of bounded variation has at most a denumerable set of points of discontinuity only of the first order*, and

(f) *A function of bounded variation may be given in the form of the sum of a continuous function with bounded variation and a jump function.*

**4.** Let us examine the function $f(x)$, which is defined on the segment $[a, b]$. The function $f(x)$ is said to be *absolutely continuous* on the segment if for any $\varepsilon > 0$ we can find $\delta > 0$ such that, for any finite system of mutually disjoint intervals, the sum of whose lengths is less than $\delta$, the sum of the absolute magnitudes of the increments of the function on these intervals is less than $\varepsilon$, i.e. for a system $\{(a_k, b_k)\}$, for which $\sum_{k=1}^{n} (b_k - a_k) < \delta$ the following condition holds

$$\sum_{k=1}^{n} |f(b_k) - f(a_k)| < \varepsilon.$$

In the definition of absolute continuity we may replace the sum of the absolute magnitudes of the increments of the function by the sum of increments, i.e. the following equality must be satisfied

$$\left| \sum_{k=1}^{n} \{f(b_k) - f(a_k)\} \right| < \varepsilon.$$

Since, in particular, instead of the system of intervals we may take just one, a function, which is absolutely continuous on the segment, is uniformly continuous on it, and hence also continuous at each of its points. As we shall show below the converse inference is not true.

A simple example of an absolutely continuous function is a function which satisfies a Lipschitz condition.

*The sum, difference and product of absolutely continuous functions are absolutely continuous functions. The quotient of absolutely continuous functions is absolutely continuous on any segment, on which the denominator is not reduced to zero.*

*An absolutely continuous function has bounded variation.* From this it immediately follows that a continuous function, whose variation is not bounded, e.g. the function examined in § 3, para. 3, $y = x\cos\frac{\pi}{2x}$, is not absolutely continuous. Even a continuous function, however, with bounded variation may not be absolutely continuous.

*In order that a function, which is continuous on the segment* $[a, b]$ *and has bounded variation on this segment, may be absolutely continuous on it, it is necessary and sufficient that the image* $f(E)$ *of any set* $E$ *of meaure zero should also be of measure zero.*†

The superposition‡ of two absolutely continuous functions may or may not be absolutely continuous. *An absolutely continuous function* $F$ *of an absolutely continuous and monotonic function* $f$, *i.e.* $F[f(x)]$, however, *is an absolutely continuous function.* Also *if* $f(x)$ *is absolutely continuous on the segment* $[a, b]$, *and* $F(y)$ *satisfies a Lipschitz condition on the segment, containing all values of* $f(x)$, *then* $F[f(x)]$ *is absolutely continuous on* $[a, b]$.

Here we have two sufficient conditions, for which the superposition of absolutely continuous functions is absolutely continuous. A more general condition is given by

G. M. Fikhtengolt's Theorem. *If a function* $f(x)$ *is absolutely continuous on a segment* $[a, b]$ *and* $F(y)$ *is absolutely continuous on the segment which contains all values of* $f(x)$, *then in order that the superposition* $F[f(x)]$ *may be absolutely continuous it is necessary and sufficient that it should be a function of bounded variation.*

**5.** Let a function $f(x)$ be defined on a set $E$. We shall represent the set of all those points $x$ of $E$, for which the inequality $f(x) > a$ holds, by $E\{f > a\}$. In some cases, in order to stress the fact that the statement is about a set of points $x$, we write $E\{x|f(x) > a\}$. Similarly we define the symbols $E\{f \leqslant a\}$, $E\{f = a\}$, etc.

The function $f(x)$ is said to be *measurable* on the set $E$ if $E$ is measurable and for any $a$ the set $E\{f > a\}$ is also measurable.

It is clear that a function, given on a set of measure zero, is always measurable, since any part of a set of measure zero is also a set of measure zero.

† N. N. Luzin called this property of a function the *N-property.*

‡ The function $y = f[\varphi(x)]$ is called the *superposition* of the functions $u = \varphi(x)$ and $y = f(u)$.

We shall describe as *a characteristic function* of a set $E \subset [a, b]$, a function $\psi(x)$ such that it is defined on $[a, b]$ and equal to unity at points of the set and zero at points of its complement. *The set $E$ and its characteristic function taken together are both measurable or not measurable.*

A function $g(x)$ is said to be *equivalent* to the function $f(x)$, if the set $E\{f \neq g\}$ has measure zero. Using the terminology introduced in § 2, we may say that functions are equivalent if they coincide almost everywhere, or alternately, if their difference is almost everywhere zero. It is easy to show that *if the function $f(x)$ is measurable, then all functions which are equivalent to $f(x)$ are also measurable.*

For a measurable function $f(x)$ with any $a$, not only the set $E\{f > a\}$, mentioned in the definition of measurability, is measurable, but also each of the sets $E\{f \geqslant a\}$, $E\{f < a\}$, $E\{f \leqslant a\}$ and $E\{f = a\}$. On the other hand, the measurability of at least one of the sets $E\{f \geqslant a\}$, $E\{f < a\}$, and $E\{f \leqslant a\}$, for any $a$, already implies the measurability of the function.

It is not difficult to establish the fact *that a function $f(x)$, defined and continuous on a segment, is measurable on it.* Also it is easy to prove that *arithmetical operations on measurable functions result only in measurable functions.*

The structure of measurable functions is made clear by the following theorems:

(a) *For any function $f(x)$ measurable and finite almost everywhere on a set $E$ and a number $\varepsilon > 0$, there exists a bounded measurable function $\varphi(x)$, different from $f(x)$ only at points of the set whose measure is less than $\varepsilon$,*

$$mE\{f \neq \varphi\} < \varepsilon.$$

(b) BOREL'S THEOREM. *Let a measurable and almost everywhere finite function $f(x)$ be given on a segment $[a, b]$. For any $\sigma > 0$ and $\varepsilon > 0$, there exists a continuous function $\psi(x)$, different from $f(x)$ by not less than $\sigma$ only at points of the set whose measure is less than $\varepsilon$,*

$$mE\{|f - \psi| \geqslant \sigma\} < \varepsilon.$$

The nature of a measurable function can be explained in more detail by the idea of a $C$-property, introduced by N. N. Luzin. Let a measurable function $f(x)$ be given on a segment $[a, b]$. We

say that $f(x)$ possesses a *C-property*, if for any $\varepsilon > 0$ we can find a perfect set $P$, on which $f(x)$ is continuous, and whose measure differs from the length of the segment by less than the given $\varepsilon$, i.e. $mP > b - a - \varepsilon$. The name $C$-property comes from the French word *continu*, which means continuous.

N. N. Luzin's Theorem. *In order that a function may possess a C-property it is necessary and sufficient that it should be measurable and finite almost everywhere.*

Thus any function, which is measurable and finite almost everywhere, may be obtained from some continuous function by its deformation on a set, whose measure is as small as we please.

The theorem of N. N. Luzin just quoted is one of the basic theorems of the theory of functions of a real variable.

## § 4. THE DERIVATIVE AND ITS GENERALIZATION

**1.** Let the function $f(x)$ be defined on a segment $[a, b]$ and be continuous at a point $x_0 \in [a, b]$. The ratio

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{f(x) - f(x_0)}{x - x_0}$$

is called the *difference ratio* for the function $f(x)$ at the given point. *The derivative* of $f(x)$ at the point $x_0$ is the limit of this difference ratio

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

if it exists. If the limit is finite we say that $f(x)$ is *differentiable* at the point $x_0$.

The continuity of the function $f(x)$ at the point $x_0$ is necessary, but not sufficient for differentiability. To be convinced of this it is sufficient to consider the behaviour of the functions $y = |x|$ and

$$y = \begin{cases} x \sin \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

at the point $x_0 = 0$. For the first of these the difference ratio is equal to $+1$ for $h > 0$ and $-1$ for $h < 0$ (Fig. 5). For the second (Fig. 6) the difference ratio in any neighbourhood of the point $x_0 = 0$ takes on all values between $-1$ and $+1$. In fact the

difference ratio is equal to

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{f(h) - f(0)}{h} = \frac{h \sin \frac{1}{h}}{h} = \sin \frac{1}{h},$$

so that all values from $-1$ to $+1$ are assumed at each part of the form

$$\frac{2}{\pi(2n - 1)} \geqslant h \geqslant \frac{2}{\pi(2n + 1)} \quad (n = 0, 1, 2, \ldots).$$

In the given examples the derivative is non-existent only at one point. Examples exist, however, of continuous functions which

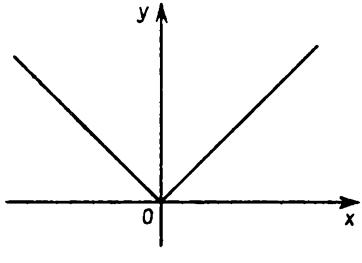


FIG. 5

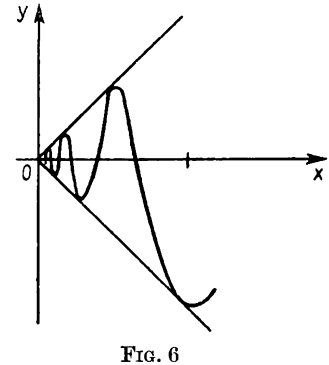


FIG. 6

have no derivative at any point of the segment. One of these will be considered in § 6 (see page 54).

If the function $f(x)$ is differentiable at each point of a segment, the function $f'(x)$ is said to be the *point derivative*. An important property of the point derivative on a segment is established by the following theorem.

DARBOUX'S THEOREM. *The point derivative of a function $f(x)$ on a segment $[a, b]$ takes all intermediate values between $f'(a)$ and $f'(b)$ on $(a, b)$. Hence it follows that the point derivative cannot have points of discontinuity of the first kind.*

**2.** If the sets of points of existence of the derivative has complete measure, we say that the derivative exists *almost everywhere*. The simplest sufficient condition for the existence of a derivative almost everywhere is the monotonicity of the function. Here the

following theorem comes in: *a function $f(x)$, which is monotonic on a segment $[a, b]$, has a finite derivative $f'(x)$ almost everywhere on the segment.* (We do not in any way assume the continuity of $f(x)$.) Hence from the theorem of § 3 (see paras. 3 and 4) it follows that

(a) *a function with bounded variation on the segment $[a, b]$ has a derivative almost everywhere on this segment*;

(b) *if a function $f(x)$ satisfies a Lipschitz condition on a segment it has a finite derivative almost everywhere*;

(c) *a function which is absolutely continuous on a segment $[a, b]$ has a finite derivative almost everywhere on it.*

In addition,

(d) *there exists no continuous function, whose derivative is infinite on a set of positive measure.*

Also *for any set $E$ of measure zero on the segment $[a, b]$ there exists a continuous increasing function $\sigma(x)$, whose derivative at each point $x \in E$ is equal to $+\infty$.*

3. Suppose that at a point $x_0$ the function $f(x)$, defined on $[a, b]$, has no derivative. If the difference ratio $\dfrac{f(x_0 + h) - f(x_0)}{h}$ has one-sided limits on the left and on the right we call them respectively the *left-hand and right-hand derivatives* at the point $x_0$, or the *derivative on the left* and *derivative on the right.* Derivatives on the left and right at the point $x_0$ are often denoted by $D_s f(x_0)$ and $D_d f(x_0)$.

In order that the function $f(x)$ may be differentiable at the point $x_0$, it is necessary and sufficient that both one-sided derivatives should exist and coincide with each other. It may happen, however, that the derivatives on the left and right exist and are different. For example, the function $f(x) = |x|$ (see Fig. 5) has no derivative at $x_0 = 0$, but $D_s f(0) = -1$, $D_d f(0) = +1$. The points at which $D_s f(x_0) \neq D_d f(x_0)$ are *angular points.*

*A set of points, at which the right-hand and left-hand derivatives exist and are different, is at most denumerable.* To use different terms, *a set of angular points is only finite or denumerable.*

At any point $x_0$ there exist an *upper derivative* $f^+(x_0)$ and a *lower derivative* $f^-(x_0)$. The upper derivative is the *upper limit of the difference ratio,*

$$f^+(x_0) = \limsup_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Similarly, the lower derivative is defined by the equation

$$f^-(x_0) = \liminf_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Thus the function

$$f(x) = \begin{cases} x \sin \dfrac{1}{x} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

has $f^+(x_0) = +1$ and $f^-(x_0) = -1$ at the point $x_0 = 0$ (see Fig. 6). In this case the slope of the straight line joining the point $(x_0, f(x_0))$ to the point $(x_0 + h, f(x_0 + h))$ varies from the bisector of the second to the bisector of the first coordinate angle and vice versa.

*Derivative numbers* are a further extension of the idea of a derivative. The number $\lambda$ (finite or infinite) is said to be *the derivative number of a function $f(x)$ at the point $x_0$*, if there exists a sequence $\{h_n\}$, tending to zero ($h_n \neq 0$), for which

$$\lim_{n \to \infty} \frac{f(x_0 + h_n) - f(x_0)}{h_n} = \lambda.$$

We shall represent the derivative number by $\lambda = Df(x_0)$. It is obvious that:

(a) *for each function $f(x)$, given on a segment $[a, b]$ there exist derivative numbers at any point*;

(b) *for the existence of a derivative $f'(x_0)$ at the point $x_0 \in [a, b]$ it is necessary and sufficient that all derivative numbers of the function at the given point should coincide.*

In many cases derivative numbers possess properties inherent in the derivative. For instance, *if a function $f(x)$ increases (decreases) on the segment $[a, b]$, then all its derivative numbers are non-negative (non-positive).*

*Dini derivative numbers* are defined in the following way. The *upper* (lower) *right-hand* derivative number is the upper (lower) limit of the difference ratio as $h$ tends to zero from the right

$$\Lambda^+ = \limsup_{h \to +0} \frac{f(x_0 + h) - f(x_0)}{h},$$

$$\lambda^+ = \liminf_{h \to +0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Similarly we define the upper and lower *left-hand* derivative numbers $\Lambda^-$ and $\lambda^-$ as

$$\Lambda^- = \limsup_{h \to -0} \frac{f(x_0 + h) - f(x_0)}{h}, \quad \lambda^- = \liminf_{h \to -0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

It is clear that for any finite function all four derivative numbers exist at each point and may be considered as functions of the same independent variable.

DINI'S THEOREM. *If any one of the Dini derivative numbers of a continuous function is continuous at a point, then at that point all the remaining derivative numbers are continuous and all four derivative numbers at the given point coincide, i.e. the function is differentiable at the point.*

The Dini derivative numbers on a set of complete measure satisfy the relations established by the following theorems.

N. N. LUZIN'S THEOREM. *If all four Dini derivative numbers of a function $f(x)$ are finite at each point of a set of positive measure, then the function is differentiable almost everywhere on that set.*

A. DENJOY'S THEOREM. *Let $f(x)$ be a finite function, defined on a set $[a, b]$. Then at each point of any set of complete measure one of the following relations holds*:

(a) $\Lambda^+ = \Lambda^- = +\infty, \quad \lambda^+ = \lambda^- = -\infty;$

(b) $\Lambda^+ = \lambda^- \neq \infty, \quad \lambda^+ = -\infty, \quad \Lambda^- = +\infty;$

(c) $\lambda^+ = \Lambda^- \neq \infty, \quad \Lambda^+ = +\infty, \quad \lambda^- = -\infty;$

(d) $\Lambda^+ = \Lambda^- = \lambda^+ = \lambda^- \neq \infty.$

**4.** A function is said to be *asymptotically continuous* at a point $x_0$, which is not an end of the segment $[a, b]$, if there exists a set $E \subset [a, b]$, for which the point $x_0 \in E$ is a density point, and the function $f(x)$ is continuous with respect to the set $E$ at the point $x_0$. Similarly, we define an asymptotic derivative: a function $f(x)$ is said to be *asymptotically differentiable* at the point $x_0$ if there exists a set $E$, for which the point $x_0$ is a density point and which is such that the difference ratio $\dfrac{f(x) - f(x_0)}{x - x_0}$ tends to a definite limit, when the point $x$ tends to $x_0$ on the set $E$. The limit so obtained is called an *asymptotic* (sometimes *approximate*) *derivative* and is denoted by $f^{[1]}(x)$.

An asymptotic derivative is an extension of the ordinary derivative, since there exist examples of continuous functions, which

have no ordinary derivatives whatever at any point of a segment, but are asymptotically differentiable almost everywhere.

*Any measurable function is asymptotically continuous almost everywhere and also, conversely, any function, which is asymptotically continous almost everywhere, is measurable.* An asymptotic derivative, however, *can* not exist on a set of positive measure.

At points, where no asymptotic derivative exists, we may consider asymptotic derivative numbers, similar to the Dini derivative numbers. A theorem similar to Denjoy's theorem holds for them.

Asymptotic derivatives of higher orders are defined by ordinary induction.

In the theory of series, along with asymptotic derivatives we use the idea of a symmetric derivative, which also extends the idea of an ordinary derivative. *A symmetric derivative* or *Schwartzian derivative* is the limit of the ratio $\frac{f(x_0+h)-f(x_0-h)}{2h}$ as $h \to 0$. The second symmetric derivative may be defined directly as the limit as $h \to 0$ of the ratio $\frac{f(x_0+h)+f(x_0-h)-2f(x_0)}{h^2}$, the limit being considered both in the ordinary and in the asymptotic sense. A symmetric derivative has a number of properties similar to those of an ordinary derivative.

## § 5. THE INTEGRATION OF FUNCTIONS

**1.** Let a function $f(x)$ be defined and bounded on the segment $[a, b]$. We shall divide the segment $[a, b]$ into $n$ parts by points $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ and denote by $\Delta_i$ the segment $[x_{i-1}, x_i]$ and its length $x_i - x_{i-1} = \Delta x_i$. We shall further denote by $M_i$ and $m_i$ respectively the upper and lower bounds of the function $f(x)$ on the segment $\Delta_i$. The difference $M_i - m_i$ is called the *variation* of the function on $\Delta_i$. We shall call the sums

$$S_n = \sum_{i=1}^{n} M_i \Delta x_i \quad \text{and} \quad s_n = \sum_{i=1}^{n} m_i \Delta x_i$$

respectively the *upper* and *lower Darboux sums* for the given partitioning.

As the maximum length of the parts tends to zero *the Darboux sums tend to definite limits* (which do not depend on the method of partitioning of the segment), which are called *the upper and*

*lower Darboux integrals of the function* $f(x)$ on the segment $[a, b]$,

$$\overline{I} = \lim_{\substack{n\to\infty \\ \Delta x_i \to 0}} S_n = \overline{\int_a^b} f(x)\,dx,$$

$$\underline{I} = \lim_{\substack{n\to\infty \\ \Delta x_i \to 0}} s_n = \underline{\int_a^b} f(x)\,dx.$$

This means that, for any $\varepsilon > 0$, there exists a $\delta > 0$, such that for all partitionings of the segment $[a, b]$, for which the length of the greatest part does not exceed $\delta$, the inequalities $|S_n - \overline{I}| < \varepsilon$ and $|s_n - \underline{I}| < \varepsilon$ hold.

If the upper and lower integrals of the function $f(x)$ on the segment $[a, b]$ coincide then the function $f(x)$ is said to be *integrable in the Riemann sense on the segment* $[a, b]$, and the common value of the upper and lower integrals is the *Riemann integral* of the function.

The Riemann integral has been considered in detail in the publication in the series a Mathematical Reference Library, *Mathematical analysis (differentiation and integration)*. We shall confine ourselves, therefore, to the formulation of two criteria for the Riemann integrability of a function.

RIEMANN'S THEOREM. *A bounded function is Riemann integrable on the segment* $[a, b]$ *when, and only when, for any* $\varepsilon > 0$ *and* $\delta > 0$ *the segment* $[a, b]$ *may be so divided into intervals that the sum of the lengths of those intervals, on which the variation is greater than* $\varepsilon$, *is less than* $\delta$.

LEBESGUE'S THEOREM. *In order that a bounded function may be Riemann integrable, it is necessary and sufficient that the set of its points of discontinuity should have measure zero.*

From Lebesgue's theorem it follows that the function cited in § 3, para. 1,

$$f(x) = \begin{cases} 0 & \text{for irrational } x, \\ \dfrac{1}{q} & \text{for } x = \dfrac{p}{q}, \end{cases}$$

is Riemann integrable on the segment [0, 1] (and its integral is zero) since the set of its points of discontinuity is the set of rational points of the segment, and consequently has measure zero. Conversely, the Dirichlet function defined there,

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational}, \\ 1 & \text{if } x \text{ is rational}, \end{cases}$$

is not Riemann integrable, because it is discontinuous everywhere. The non-integrability of the Dirichlet function easily follows from the definition. In fact its upper Darboux sums are always equal to unity, because, with any partitioning, at least one rational point can be found on each part. On a similar principle the lower Darboux sums for the Dirichlet function are always zero. Thus the upper and lower sums can have no common limit.

**2.** If in the Darboux sums we replace the length of the segment $\Delta x_i$ by the increment of some *integrating* (or *weight*) *function* $\varphi(x)$ on this segment, we arrive at the Darboux–Stieltjes sums,

$$S_n = \sum_{i=1}^{n} M_i[\varphi(x_i) - \varphi(x_{i-1})],$$

$$s_n = \sum_{i=1}^{n} m_i[\varphi(x_i) - \varphi(x_{i-1})].$$

The common limit of the sums, if it exists and is independent of the method of partitioning of the segment, is called the *Stieltjes integral* of the function $f(x)$ with respect to the integrating function $\varphi(x)$ on the segment $[a, b]$ and is denoted by $\int_a^b f(x)\, d\varphi(x)$. We often speak also of the *Riemann–Stieltjes integral*. It is easy to see that the Riemann integral is a particular case of the Riemann–Stieltjes integral with the integrating function $\varphi(x) \equiv x$.

*If the function* $f(x)$ *is continuous and* $\varphi(x)$ *has bounded variation on the segment* $[a, b]$, *the integral* $\int_a^b f(x)\, d\varphi(x)$ *exists.*

*If the function* $f(x)$ *is continuous and the point derivative* $\varphi'(x)$ *exists and is Riemann integrable on the segment* $[a, b]$, *then*

$$\int_a^b f(x)\, d\varphi(x) = \int_a^b f(x)\, \varphi'(x)\, dx,$$

where the integral on the left is understood in the Stieltjes sense and the integral on the right is an ordinary Riemann integral.

To calculate the Stieltjes integral we often use the formula for integration by parts, which may be given in the following form. *If one of the integrals* $\int_a^b f(x)\, d\varphi(x)$ and $\int_a^b \varphi(x)\, df(x)$ *exists then the other*

*also exists, and the following equation holds,*

$$\int_a^b f(x)\,d\varphi(x) + \int_a^b \varphi(x)\,df(x) = f(b)\,\varphi(b) - f(a)\,\varphi(a).$$

In particular, if the function $f(x)$ has an integrable point derivative, and $\varphi(x)$ is continuous,

$$\int_a^b f(x)\,d\varphi(x) = \Big[f(x)\,\varphi(x)\Big]_a^b - \int_a^b \varphi(x)\,f'(x)\,dx,$$

the integral on the right being understood as a Riemann integral.

*Change of variables* in a Stieltjes integral is possible with ordinary conditions of monotonicity of the transformation involved: *if $x = \psi(t)$ is a monotonic function, which mutually maps the segment $c \leqslant t \leqslant d$ uniquely onto the segment $a \leqslant x \leqslant b$, then*

$$\int_a^b f(x)\,d\varphi(x) = \int_c^d f[\psi(t)]\,d\varphi[\psi(t)],$$

*the existence of one integral being implied by the existence of the other.*

If the function $f(x)$ satisfies the inequality $|f(x)| \leqslant M$ for $a \leqslant x \leqslant b$ and $\varphi(x)$ is a function of bounded variation, then for the integral $\int_a^b f(x)\,d\varphi(x)$, if it exists, we have the following value

$$\left|\int_a^b f(x)\,d\varphi(x)\right| \leqslant M \bigvee_a^b(\varphi).$$

With the same conditions for $\varphi(x)$

$$\left|\int_a^b f(x)\,d\varphi(x)\right| \leqslant \int_a^b |f(x)|\,d \bigvee_a^x(\varphi).$$

The calculation of a Stieltjes integral, with an integrating function $\varphi(x)$ of bounded variation, may always be reduced to the calculation of an integral with a continuous integrating function of bounded variation. In fact, in virtue of the property (*f*), § 3, para. 3, for a function $\varphi(x)$ of bounded variation, the statement $\varphi(x) = \psi(x) + s(x)$ is possible, where $\psi(x)$ is continuous and has bounded variation and $s(x)$ is a jump function. *Therefore a Stieltjes integral permits the following canonical statement,*

$$\int_a^b f(x)\,d\varphi(x) = \int_a^b f(x)\,d\psi(x) + f(a)[\varphi(a+0) - \varphi(a)] +$$
$$+ \sum_{k=1}^{\infty} f(x_k)[\varphi(x_k+0) - \varphi(x_k-0)] + f(b)[\varphi(b) - \varphi(b-0)],$$

where $\psi(x)$ is a continuous function, and $\{x_k\}$ is a jump point of the function $\varphi(x)$. The existence of the integral on the left-hand side guarantees the convergence of the series on the right-hand side.

**3.** Let us consider a bounded measurable function $f(x)$, defined on a measurable set $E \subset [a, b]$. We choose a segment $[A, B]$ so that

$$f(E) \subset (A, B),$$

and divide it into $n$ parts by the points $A = y_0 < y_1 < y_2 < \cdots < y_n = B$. We now define the set $E_k$, on which the function $f(x)$ takes values included between $y_{k-1}$ and $y_k$,

$$E_k = E\{y_{k-1} \leqslant f(x) < y_k\}.$$

It is clear that all the sets $E_k$ $(k = 1, 2, \ldots, n)$ are measurable and mutually disjoint. In addition

$$E = \sum_{k=1}^{n} E_k, \quad mE = \sum_{k=1}^{n} mE_k.$$

Now we may construct the sums

$$s_n = \sum_{k=1}^{n} y_{k-1} mE_k, \quad S_n = \sum_{k=1}^{n} y_k mE_k$$

which we shall call *the lower and upper Lebesgue sums* respectively. If the number $n$ increases without limit, while the maximum length of the parts $(y_{k-1}, y_k)$ tends to zero, the *upper and lower Lebesgue sums tend to a common limit, which is called the Lebesgue integral of the function $f(x)$ with respect to the set $E$.* This limit does not depend on the choice of the segment $[A, B]$ nor on the method of its partitioning.

In Fig. 7 rectangles have been shaded, which together make up one of the terms of the lower Lebesgue sum for the case where the set $E$ coincides with the segment $[a, b]$. The set $E_k$ here consists of several segments.

The Lebegue integral of the function $f(x)$ with respect to the set $E$ is denoted by $\int_E f(x)\,dx$, and if the set $E$ coincides with the segment $[a, b]$ by $\int_a^b f(x)\,dx$. Sometimes in order to stress the fact that the discussion is about a Lebesgue integral, we put the sign $(L)$ before the integral.

We shall call a function, which is integrable in the Lebesgue sense, *summable*. As we have remarked above, *every measurable bounded function is summable.*

The Lebesgue integral has intrinsic properties in common with the Riemann integral.

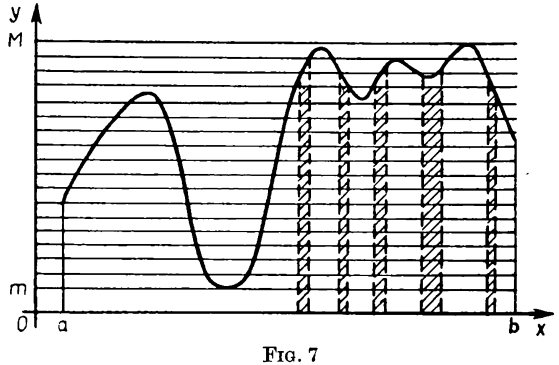


FIG. 7

(a) THE THEOREM OF THE MEAN. *If a measurable function $f(x)$ on a measurable set $E$ satisfies the inequality $m \leqslant f(x) \leqslant M$, then*

$$m \,.\, mE \leqslant \int_E f(x)\, dx \leqslant M \,.\, mE.$$

(b) THE THEOREM OF COMPLETE ADDITIVITY. *If a measurable set $E$ is divided into the sum of finite or denumerable sets of mutually disjoint measurable sets $E_k$ $(k = 1, 2, \ldots)$ and $f(x)$ is a measurable bounded function, given on $E$, then*

$$\int_E f(x)\, dx = \sum_k \int_{E_k} f(x)\, dx.$$

From these theorems it follows that

$$\int_E c\, dx = cmE;$$

*if $f(x) \geqslant 0$ on $E$, then*

$$\int_E f(x)\, dx \geqslant 0;$$

*if $mE = 0$, then for any bounded function*

$$\int_E f(x)\, dx = 0;$$

*if the functions $f(x)$ and $\varphi(x)$ coincide (are equivalent) almost everywhere on $E$, then*

$$\int_E f(x)\,dx = \int_E \varphi(x)\,dx;$$

*if $f(x) \geqslant 0$ on $E$ and $\int_E f(x)\,dx = 0$, then $f(x) \equiv 0$ almost everywhere on $E$.*

(c) *If on a measurable set $E$ we are given measurable and bounded functions $f(x)$ and $\varphi(x)$, then*

$$\int_E [f(x) \pm \varphi(x)]\,dx = \int_E f(x)\,dx \pm \int_E \varphi(x)\,dx;$$

$$\int_E cf(x)\,dx = c\int_E f(x)\,dx \quad (c \text{ is a constant});$$

*if $f(x) \geqslant \varphi(x)$ on the set $E$,*

$$\int_E f(x)\,dx \geqslant \int_E \varphi(x)\,dx.$$

(d) *If the function $f(x)$ is measurable and bounded on the set $E$, then*

$$\left|\int_E f(x)\,dx\right| \leqslant \int_E |f(x)|\,dx.$$

The Lebesgue integral represents a generalization of the Riemann integral; *if a function $f(x)$ is integrable in the Riemann sense on the segment $[a, b]$, then it is integrable also in the Lebesgue sense and the integrals coincide.* This statement makes a difference of notation for Riemann and Lebesgue integrals unnecessary in the majority of cases, viz. the special signs $(R)$ and $(L)$. On the other hand, *summable functions exist on $[a, b]$, which are not integrable in the Riemann sense.* As an example we may point to the Dirichlet function cited in § 3, para. 1. As we have shown there, the Riemann integral for this function does not exist. Nevertheless, it is summable because it is equal to zero almost everywhere, so that the Lebesgue integral of this function exists and is equal to zero.

The difference between the processes of Riemann and Lebesgue integration may be figuratively explained by the following example due to Lebesgue. Suppose that it is required to calculate the value of a pile of coins of different denominations lying on a table. The process of Riemann integration consists of taking coins from the pile in succession in the order in which they lie and adding their

values. The process of Lebesgue integration consists of previously sorting out the coins in accordance with their values, each of the possible values being multiplied by the number of respective coins.

**4.** The Lebesgue integral may also be defined for unbounded functions. Let us first consider a *non-negative*, measurable function $f(x)$, defined on a measurable set $E$. Let $N$ be some natural number. We shall define a function $f_N(x)$ in the following way:

$$f_N(x) = \begin{cases} f(x) & \text{if } f(x) \leqslant N, \\ N & \text{if } f(x) > N. \end{cases}$$

The function $f_N(x)$ is sometimes called a *cut* of the function $f(x)$ by the number $N$.

The function $f_N(x)$ is measurable and bounded on $E$ and is therefore integrable in the Lebesgue sense. In addition the sequence of integrals with respect to the cut $\left\{\int_E f_N(x)\,dx\right\}$ $(N = 1, 2, \ldots)$ increases monotonically, so that there exists a limit which is finite or equal to $+\infty$, viz.

$$\lim_{N\to+\infty} \int_E f_N(x)\,dx.$$

This limit is called *the Lebesgue integral of the non-negative function $f(x)$ on the set $E$ and is denoted by the symbol* $\int_E f(x)\,dx$. If this integral is finite the function $f(x)$ is said to be *Lebesgue integrable* or *summable* on the set $E$. From the definition it follows that *a function, which is summable on a set $E$, is almost everywhere finite on it*;

*if $mE = 0$, any non-negative function is summable on $E$ and its integral on the set $E$ is zero*;

*if $f(x)$ and $\varphi(x)$ are non-negative measurable functions on $E$ and $f(x) \geqslant \varphi(x)$, then* $\int_E f(x)\,dx \geqslant \int_E \varphi(x)\,dx$;

*if $f(x)$ is non-negative and measurable on $E$ and a subset $E_1 \subset E$ is measurable, then*

$$\int_E f(x)\,dx \geqslant \int_{E_1} f(x)\,dx;$$

*if $f(x)$ and $\varphi(x)$ are equivalent on $E$, then*

$$\int_E f(x)\,dx = \int_E \varphi(x)\,dx; \quad \text{if} \quad \int_E f(x)\,dx = 0 \quad \text{and} \quad f(x)$$

*is non-negative, then $f(x) = 0$ almost everywhere on $E$*;

*if $f(x)$ and $\varphi(x)$ are non-negative measurable functions on $E$ and $c \geqslant 0$ is a constant, then*

$$\int_E [f(x) + \varphi(x)]\,dx = \int_E f(x)\,dx + \int_E \varphi(x)\,dx;$$

$$\int_E cf(x)\,dx = c\int_E f(x)\,dx.$$

An important part is played in the theory of the Lebesgue integral by

THE THEOREM OF COMPLETE ADDITIVITY. *Let a measurable set $E$ be given in the form of the sum of a finite or denumerable set of mutually disjoint measurable sets,*

$$E = \sum_k E_k.$$

*Then for any non-negative function $f(x)$, summable on the set $E$*

$$\int_E f(x)\,dx = \sum_k \int_{E_k} f(x)\,dx.\dagger$$

We now turn to the consideration of unbounded functions which assume values of any sign. In order to define the integral of such a function $f(x)$, which is defined and measurable on a measurable set $E$, we shall introduce $f^+(x)$ and $f^-(x)$, assuming that

$$f^+(x) = \begin{cases} f(x), & \text{if } f(x) \geqslant 0, \\ 0, & \text{if } f(x) < 0; \end{cases}$$

$$f^-(x) = \begin{cases} 0, & \text{if } f(x) \geqslant 0, \\ -f(x), & \text{if } f(x) < 0. \end{cases}$$

Both functions $f^+(x)$ and $f^-(x)$ are measurable and non-negative and

$$f(x) = f^+(x) - f^-(x).$$

It is natural to call the difference $\int_E f^+(x)\,dx - \int_E f^-(x)\,dx$ the integral of the function $f(x)$. This difference, however, will make no sense in the case where both the integrals which enter into it are infinite. Therefore we call this difference the integral of an unbounded function $f(x)$, only in the case where at least one of the

† This equation must be understood in the following sense: if the right-hand side takes on a finite value or $+\infty$, then the left-hand side takes on the same value or $+\infty$, and vice versa.

integrals $\int_E f^+(x)\,dx$ and $\int_E f^-(x)\,dx$ is finite, i.e. at least one of the functions $f^+(x)$ and $f^-(x)$ is summable. Here $\int_E f(x)\,dx$ may take on both finite values and also the values $+\infty$ or $-\infty$. If both functions, $f^+(x)$ and $f^-(x)$, are summable, the function $f(x)$ is said to be *summable* (*integrable in the Lebesgue sense*) and its integral is assumed to be equal to

$$\int_E f(x)\,dx = \int_E f^+(x)\,dx - \int_E f^-(x)\,dx.$$

*A measurable function $f(x)$ is summable when, and only when, the function $|f(x)|$ is summable.* Then

$$\left|\int_E f(x)\,dx\right| \leqslant \int_E |f(x)|\,dx.$$

The majority of the properties of the integrals, which are valid for bounded or non-negative unbounded functions, can be brought under the general case. Thus, for instance,

*a summable function is finite almost everywhere*;

*a function which is summable on a set $E$ is summable also on any of its measurable subsets*;

*on a set of measure zero any function is summable and its integral is zero*;

*if functions $f(x)$ and $\varphi(x)$ are measurable on a set $E$ and $|\varphi(x)| \leqslant f(x)$, then from the summability of $f(x)$ there follows the summability of $\varphi(x)$*;

*if the functions $f(x)$ and $\varphi(x)$ are equivalent on a set $E$, then they are simultaneously summable or not; in the case of their summability their integrals on $E$ are equal to each other.*

For the case of an arbitrary summable function, the complete additivity of the Lebesgue integral, in the general form given for bounded or non-negative functions, does not hold. Here only the following theorem is true:

THE THEOREM OF FINITE ADDITIVITY OF AN INTEGRAL. *If the set $E$ is given in the form of the sum of a finite number of mutually disjoint measurable sets, $E = \sum_{k=1}^{n} E_k$, and the function $f(x)$ is summable on each of the sets $E_k$, then it is summable also on $E$ and*

$$\int_E f(x)\,dx = \sum_{k=1}^{n} \int_{E_k} f(x)\,dx.$$

Nevertheless, theorems on the complete additivity of the Lebesgue integral are also possible, but here some additional restrictions are required. The complete additivity holds, for instance, in the following cases.

(a) *If the function $f(x)$ is summable on a set $E$, which is given in the form of the sum of a denumerable set of mutually disjoint measurable sets, $E = \sum_{k=1}^{\infty} E_k$, then*

$$\int_E f(x)\,dx = \sum_{k=1}^{\infty} \int_{E_k} f(x)\,dx.$$

(b) *Let a set $E$ be given in the form of the sum of a denumerable set of mutually disjoint measurable sets, $E = \sum_{k=1}^{\infty} E_k$, and the function $f(x)$ be summable on each $E_k$, then if $\sum_{k=1}^{\infty} \int_{E_k} |f(x)|\,dx < \infty$, $f(x)$ is summable on $E$ and*

$$\int_E f(x)\,dx = \sum_{k=1}^{\infty} \int_{E_k} f(x)\,dx.$$

For summable functions the following two properties are also true:

*If $f(x)$ is summable on $E$ and $k$ is a constant, the function $kf(x)$ is summable on $E$ and*

$$\int_E kf(x)\,dx = k\int_E f(x)\,dx.$$

*If $f(x)$ and $\varphi(x)$ are summable on $E$, then their sum and difference are also summable on $E$ and*

$$\int_E [f(x) \pm \varphi(x)]\,dx = \int_E f(x)\,dx \pm \int_E \varphi(x)\,dx.$$

A special part is played in what follows by the property of the *absolute continuity* of the Lebesgue integral, which is established in the following theorem.

THE THEOREM OF ABSOLUTE CONTINUITY OF THE LEBESGUE INTEGRAL. *Let the function $f(x)$ be summable on a measurable set $E$. Then for any $\varepsilon > 0$ we can find $\delta > 0$ such that, for all sets $e \subset E$ with $me < \delta$, we have*

$$\left| \int_e f(x)\,dx \right| < \varepsilon.$$

**5.** Suppose that a summable function $f(x)$ is given on a segment $[a, b]$. *The indefinite Lebesgue integral* of the function $f(x)$ is any function

$$F(x) = \int_a^x f(t)\,dt + C,$$

where $C$ is any constant and the integral is taken in the Lebesgue sense. *The indefinite Lebesgue integral is an absolutely continuous function,* which follows from the theorem of absolute continuity of the integral, cited in § 5, para. 4. In virtue of the theorem, § 4, para. 2, the indefinite Lebesgue integral has a finite derivative almost everywhere, which is summable. Moreover, *the derivative of an indefinite Lebesgue integral* $F(x) = \int_a^x f(t)\,dt$ *is equal to the function* $f(x)$ *at almost every point* $x \in [a, b]$.

On the other hand, *any absolutely continuous function is the indefinite Lebesgue integral of its own derivative,* i.e. for an absolutely continuous function the *Newton–Leibniz formula*

$$\int_a^x f'(t)\,dt = f(x) - f(a)$$

is correct.

We must note that the formula may prove to be false for the case of a point derivative $f'(x)$, because a point derivative is not necessarily a summable function. Thus for the function

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x^2}, & \text{for } x > 0, \\ 0, & \text{for } x = 0, \end{cases}$$

the point derivative exists, but is not summable on any segment which contains the point $x = 0$.

For the correctness of the Newton–Leibniz formula it is sufficient to superimpose on the point derivative some additional limitations. Thus, *if the derivative* $f'(x)$ *exists at each point and is bounded on the segment, then*

$$\int_a^x f'(t)\,dt = f(x) - f(a).$$

This formula is correct only when $f'(x)$ is everywhere finite and summable. Hence, in particular, it follows that it is correct when $f(x)$ is a function of bounded variation and its derivative is everywhere finite.

If the derivative exists only almost everywhere then, generally speaking, no additional conditions imposed on the derivative permit us to take as correct the Newton–Leibniz formula for all $x$. We can actually construct an example of a non-decreasing continuous function, which is not equal to zero, whose derivative exists and is zero almost everywhere. For this let us consider the Cantor perfect set constructed in § 1, para. 8. Let $x$ be a point of the Cantor set and $x = 0 \cdot a_1 a_2 a_3 \ldots$ its *ternary* representation. It follows from the construction that the elements of the expansion $a_1, a_2 \ldots$ may take only the values 0 and 2. Therefore we may assume that $b_n = \frac{1}{2}a_n$ and define the value of the function by means of the

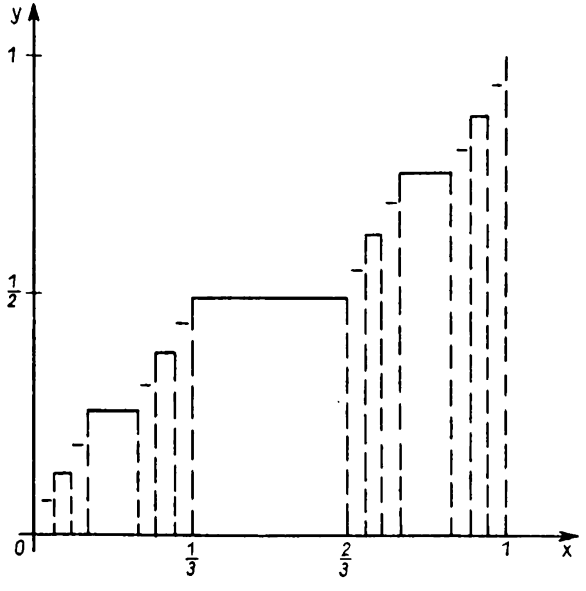


FIG. 8

*binary* notation $f(x) = 0 \cdot b_1 b_2 b_3 \ldots$ It remains to determine the function $f(x)$ in adjacent intervals, where we assume that it is constant. The latter is possible because the value of the function $f(x)$ at the ends of an adjacent interval is given by the binary expansions $0 \cdot b_1 b_2 b_3 \cdots b_m 0111 \ldots$ and $0 \cdot b_1 b_2 \cdots b_m 1000 \ldots$ which are equal to one and the same number. The graph of the function $f(x)$ is drawn in Fig. 8.

It is easy to prove that f($x$) is a non-decreasing and continuous function. Its derivative exists everywhere, except for points of the Cantor set, i.e. almost everywhere on [0, 1].

However

$$\int_0^1 f'(t)\,dt \neq f(1) - f(0),$$

because $f(1) = 1, f(0) = 0$ and $\int_0^1 f'(t)\,dt = 0$, since $f'(t) = 0$ almost everywhere. Thus the function $f(t)$ is not equal to the integral of its own derivative. Hence among other things it follows that $f(x)$ is not an absolutely continuous function.

The Lebesgue integral of the derivative may be used to find the total variation of an absolutely continuous function. For instance if $f(x)$ is summable on $[a, b]$ and

$$F(x) = \int_a^x f(t)\,dt,$$

then

$$\bigvee_a^b (F) = \int_a^b |f(t)|\,dt,$$

i.e. *the total variation of an absolutely continuous function is equal to the integral of the modulus of its derivative.*

The general formulae for the change of variable and integration by parts are correct for a Lebesgue integral with some additional conditions.

If $f(x)$ is summable on the segment $[a, b]$ and $x = \varphi(t)$ assumes values only from $[a, b]$, then

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f[\varphi(t)]\,\varphi'(t)\,dt,$$

where $a = \varphi(\alpha)$, $b = \varphi(\beta)$. This formula holds, for example, for the case *where an absolutely continuous function $\varphi(t)$ is strictly monotonic on* $[\alpha, \beta]$.

If $f(x)$ is absolutely continuous, and $\varphi(x)$ is summable on $[a, b]$ and $\Phi(x)$ denotes the indefinite integral of $\varphi(x)$, the formula *for integration by parts* is correct, i.e.

$$\int_a^b f(x)\,\varphi(x)\,dx = \Big[f(x)\,\Phi(x)\Big]_a^b - \int_a^b f'(x)\,\Phi(x)\,dx.$$

**6.** If the function $\varphi(x)$ vanishes outside the segment $[a, b]$, together with its primitive $\Phi(x)$ (including the ends of the segment), and the function $f(x)$ is differentiable, the integrated term vanishes and the formula for integration by parts becomes

$$\int_a^b f(x)\,\varphi(x)\,dx = -\int_a^b f'(x)\,\Phi(x)\,dx$$

or alternatively,

$$\int_a^b f(x)\,\Phi'(x)\,dx = -\int_a^b f'(x)\,\Phi(x)\,dx.$$

This formula was put forward by S. L. Sobolev as the basis of a definition of a generalized derivative which was then used in the solution of a series of problems in mathematical physics. For instance, the function $\omega(x)$ is said to be a *generalized derivative of* $f(x)$ *in the Sobolev sense* if, for any function $\Phi(x)$, infinitely differentiable and vanishing outside some segment $[a, b]$, the following equation holds,

$$\int_a^b f(x)\,\Phi'(x)\,dx = -\int_a^b \omega(x)\,\Phi(x)\,dx.$$

For more details of generalized differentiation and other operations on *generalized functions* see a future publication in the Mathematical Reference Library.

**7.** The above definitions of Riemann, Stieltjes and Lebesgue integrals are *constructive*, i.e. they indicate operations which must be performed for finding these integrals.

In contrast to this a *descriptive* definition indicates only the properties which the entity defined must possess. Thus, the definition of a primitive as a function, whose derivative is a given function, is descriptive. A descriptive definition is also said to be *axiomatic*.

The Lebesgue integral of a bounded measurable function may be defined descriptively as follows.

*Let us set up a correspondence between each bounded, measurable function* $f(x)$, *defined on a finite segment* $[a, b]$, *and some finite number* $\int_a^b f(x)\,dx$, *which we shall call the integral of* $f(x)$ *on* $[a, b]$, *and*

*which possesses the following properties*:

(1) *For any a, b and h*

$$\int_a^b f(x)\,dx = \int_{a+h}^{b+h} f(x-h)\,dx.$$

(2) *For any a, b and c*

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx + \int_c^a f(x)\,dx = 0.$$

(3) $$\int_a^b [f(x) + \varphi(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b \varphi(x)\,dx.$$

(4) *If* $f(x) \geqslant 0$ *and* $b > a$

$$\int_a^b f(x)\,dx \geqslant 0.$$

(5) $$\int_0^1 1\,dx = 1.$$

(6) *If, with increase of n, the bounded measurable functions* $f_n(x)$, *as they increase, approach some bounded measurable function* $f(x)$, *then the integral of* $f_n(x)$ *approaches the integral of* $f(x)$.

It is clear that the Lebesgue integral, which was defined constructively in § 5, para. 3, has the above six† properties. It is unusual, however, for the converse to be true. For instance, as Lebesgue pointed out, *if the integral possesses the properties* 1–6, *then it may be obtained by the construction described in* § 5, para. 3. *Therefore properties* 1-6 *are said to be a descriptive definition of a Lebesgue integral.*‡

The possibility of a descriptive definition of an integral shows that the Lebesgue integral is, so to speak, the more natural and reasonable integral for a function of one variable. Further

† The possession of the properties 1–5 follows from the theorems in § 5, paras. 3 and 4. As for property 6, it depends on the possibility of *passage to a limit under the integral sign* and will be examined in detail in § 6.

‡ As N. N. Luzin showed later, the formulation of Lebesgue's axioms requires more precise and additional terms; even so these axioms in their more precise form do not completely cover the class of summable functions.

extensions of the idea of an integral, which have to be investigated in connection with other particular problems, are not of such a general character.

8. The idea of an integral can be extended to the case of unbounded sets. Let $E$ be an arbitrary set lying on a straight line. We shall represent by $E_a$ the intersection of the set $E$ with the interval $(-a, a)$, $E_a = E \cap (-a, a)$. Let us assume that all sets $E_a (a > 0)$ are measurable. The function $f(x)$ is said to be *summable on* $E$ if it is summable on each set $E_a$ and there exists a finite limit $\lim_{a\to\infty} \int_{E_a} |f(x)|\, dx$. For a summable function it is assumed that

$$\int_E f(x)\, dx = \lim_{a\to\infty} \int_{E_a} f(x)\, dx.$$

From this definition it follows that for an integral on unbounded sets the basic properties of the Lebesgue integral apply.

## § 6. SEQUENCES OF FUNCTIONS

1. Let $\{f_n(x)\}$ be a functional sequence, all of whose terms are defined on one and the same set $E$. We shall consider from now on only bounded sets $E$. We shall say that a sequence *converges at a point* $x_0 \in E$ if the numerical sequence $\{f_n(x_0)\}$ converges. The point $x_0$ is called *the point of convergence* of the given sequence of functions. The set $E_1 \subset E$ of all points of convergence will be called the *set of convergence* of the sequence $\{f_n(x)\}$. If $E_1 = E$, i.e. each point of $E$ is a point of convergence, we say that the functional sequence *converges on* $E$ *at each point* (or *converges everywhere on* $E$).

Assuming, at each point $x_0 \in E_1$, that

$$f(x_0) = \lim_{n\to\infty} f_n(x_0),$$

we note that the given functional sequence defines, at each point of $E_1$, *a limit function*, to which the sequence converges on the set $E_1$ (or *at each point of* $E$, if $E_1 = E$).

If the difference $E - E_1$ is a set of measure zero, then it is customary to say that *the sequence* $\{f_n(x)\}$ *converges to the function* $f(x)$ *almost everywhere on* $E$. We write

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{almost everywhere on} \quad E.$$

Together with the sequence $\{f_n(x)\}$ we often consider the *series of functions*

$$u_1(x) + u_2(x) + \cdots + u_n(x) + \cdots$$

or briefly, $\sum_{n=1}^{\infty} u_n(x)$. Corresponding to this series we have the sequence of its *partial sums* $\{s_n(x)\}$, whose terms are defined by the equations

$$s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x) \quad (n = 1, 2, \ldots).$$

The series is said to be *convergent* at each point or almost everywhere, if the sequence of its partial sums converges.

Any sequence may be considered as a sequence of partial sums of some series. In actual fact, for the sequence $\{f_n(x)\}$ such a series is

$$f_1(x) + (f_2(x) - f_1(x)) + \cdots + (f_n(x) - f_{n-1}(x)) + \cdots = \\ = f_1(x) + \sum_{n=1}^{\infty} (f_{n+1}(x) - f_n(x)).$$

*If a sequence of measurable functions converges everywhere* (or *almost everywhere*) *on a segment, then the limit function is also measurable on this segment.* It is clear that this is true also for a series.

**2.** The convergence of a sequence $\{f_n(x)\}$ to a function $f(x)$ at each point of the set $E$ means that for any $\varepsilon > 0$ and any $x \in E$ there can be found such a number $N$, *depending on* $\varepsilon$ *and on* $x$, $N = N(\varepsilon, x)$, such that for all $n > N$

$$|f_n(x) - f(x)| < \varepsilon.$$

We say that the sequence converges *uniformly* when it is possible to choose one and the same number $N$ for all $x \in E$. In more detail; a sequence $\{f_n(x)\}$ *converges uniformly to a function* $f(x)$ *on the set* $E$, if for any $\varepsilon > 0$ we can find $N$, depending only on $\varepsilon$, $N = N(\varepsilon)$, such that for all $n > \mathrm{N}$ and $x \in E$

$$|f_n(x) - f(x)| < \varepsilon.$$

From the uniform convergence of a sequence on a set $E$ its convergence at each point of this set follows, but the converse is not true.

A sequence of continuous functions, which converges at each point, may have as its limit a discontinuous function. Thus, the sequence $\{x^n\}$ converges at each point of the segment $[0, 1]$ to the following function:

$$f(x) = \begin{cases} 0, & \text{if } 0 \leqslant x < 1, \\ 1, & \text{if } x = 1, \end{cases}$$

which is discontinuous for $x = 1$. *If a sequence of continuous functions* $\{f_n(x)\}$ *converges uniformly on a set* $E$, *however, the limit function* $f(x)$ *is continuous on* $E$.

The following question provides great interest. Let $\lim_{n\to\infty} f_n(x) = f(x)$ on a set $E$. Does the following equation hold here,

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx\,?$$

If the equation holds, we say that it *allows a passage to the limit under the integral sign*. The integral may be taken in various senses.

For a Riemann integral, with convergence of a sequence at each point of the segment, the passage to the limit is not always permitted, because the limit function may prove to be unintegrable. Nevertheless, *if a sequence* $\{f_n(x)\}$ *of functions, which are Riemann integrable on the segment* $[a, b]$, *converges uniformly, the limit function* $f(x)$ *is also Riemann integrable and the passage to the limit under the integrable sign is possible.* A similar statement is true also for a Lebesgue integral, but for this we have a more general theorem, about which we shall speak below.

The situation is different with the differentiation of a sequence. *Let* $\{f_n(x)\}$ *be a sequence of functions, which are differentiable on a segment* $[a, b]$ and converge at at least one point on it. *Then if the sequence of derivatives* $\{f_n'(x)\}$ *converges uniformly on* $[a, b]$ *to a function* $\varphi(x)$, *the original sequence converges uniformly on* $[a, b]$ *to a differentiable function, whose derivative is equal to* $\varphi(x)$.

Uniform convergence of a sequence of differentiable functions is not itself sufficient for the limit function to be differentiable. On this we usually base the examples of continuous functions, which have no derivative (see § 4, para. 1), the first of which was constructed by K. Weierstrass. Let us cite one such example.

VAN DER WAERDEN'S EXAMPLE. Let us define for each $n = 1, 2, \ldots$ and for any $x \in [0, 1]$ a function $u_n(x)$ as the distance from the point $x$ to its nearest point of the form $\frac{m}{4^n}$, where $m$ is a whole number (parts of the graphs of functions $u_1(x)$, $u_2(x)$, $u_3(x)$ and $u_4(x)$ are shown in Fig. 9). It is clear that each of the functions $u_n(x)$ is continuous. Now let $f(x) = \sum_{n=1}^{\infty} u_n(x)$. This series converges uniformly, so that everywhere on $[0, 1]$, $|u_n(x)| < \frac{1}{4^n}$. Therefore

the function $f(x)$ is also continuous. In addition it has no derivative at any one inner point of the interval (0, 1).

Now let $x = \sum_{k=1}^{\infty} \frac{a_k}{4^k}$ be a point of the interval. Let us define the sequence $\{x_k\}$, which converges to $x$, assuming that

$$x_k = x + (-1)^{a_k} \frac{1}{4^k}.$$

Now if $k > n$, the points $x$ and $x_k$ have a nearest common point of the form $m/4^n$, and are situated on the same side of it. Hence it follows that in this case $u_n(x_k) - u_n(x) = \pm(x_k - x)$. If $k \leqslant n$, then $u_n(x_k) - u_n(x) = 0$. Therefore with $k$ a constant

$$f(x_k) - f(x) = \sum_{n=1}^{k-1} \pm (x_k - x) = p(x_k - x),$$

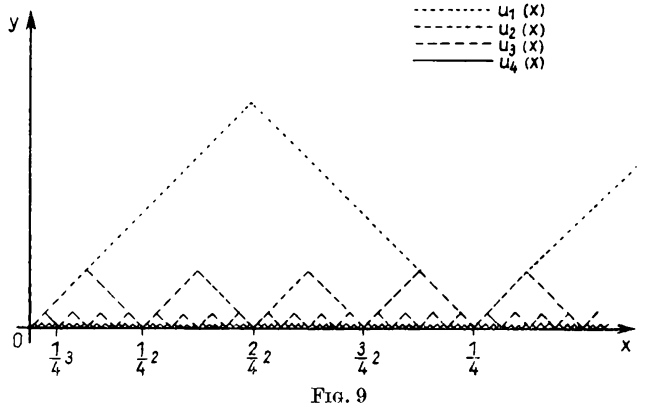


FIG. 9

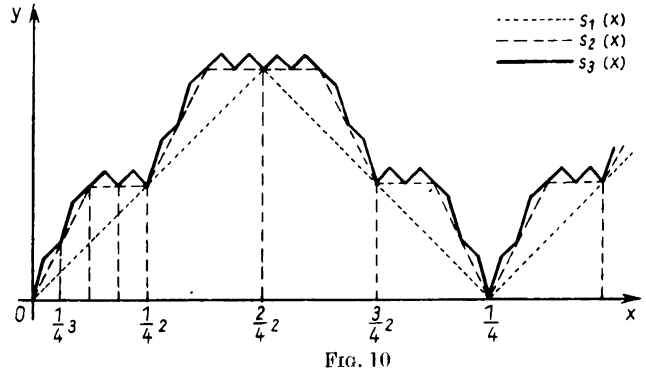


FIG. 10

where $p$ is a whole number even or odd with $k - 1$. Thus the difference ratio

$$\frac{f(x_k) - f(x)}{x_k - x}$$

as $k \to \infty$ can have no limit, because the function $f(x)$ is not differentiable. In Fig. 10 we show part of the graph of a few partial sums of the series, which determines the function $f(x)$.

**3.** From the fact that a sequence of continuous functions converges to a continuous function, it does not follow that the convergence is uniform. This conclusion is possible only with some additional assumptions. For instance, *if a sequence of continuous functions converges to a continuous limit, as it increases, then the convergence is uniform.*

The following theorem shows that in some sense uniform convergence occurs in any case.

D. F. EGOROV'S THEOREM. *Let a sequence of measurable and almost everywhere finite functions* $\{f_n(x)\}$ *converge almost everywhere on the segment* $[a, b]$. *Then for any* $\varepsilon > 0$ *there exists a set* $E \subset [a, b]$, *with measure* $mE > b - a - \varepsilon$, *on which the sequence* $\{f_n(x)\}$ *converges uniformly.*

Egorov's theorem has a wide application in the theory of functions and together with Luzin's theorem on the $C$-property (§ 3, para. 5) holds a central position in the theory of functions: *except for a set of as small measure as we please any function is continuous and any convergence is uniform.*

The possibility of defining a function by means of a series of polynomials is established by the following theorems.

K. WEIERSTRASS' THEOREM. *Any continuous function* $f(x)$ *on a segment* $[a, b]$ *is the sum of a series of polynomials, which converges to* $f(x)$ *absolutely and uniformly on* $[a, b]$.

N. N. LUZIN'S THEOREM. *For any function* $f(x)$, *which is measurable on the segment* $[a, b]$, *there exists a series of polynomials which converges absolutely to* $f(x)$ *almost everywhere on* $[a, b]$.

**4.** Let $\{f_n(x)\}$ be a sequence of functions, defined on a set $E \subseteq [a, b]$. We say that this sequence *converges in the mean to the function* $f(x)$ if

$$\int_E [f_n(x) - f(x)]^2 \, dx \to 0 \quad \text{as} \quad n \to \infty,$$

where the integral is taken in the Lebesgue sense. Sometimes we speak more precisely of *convergence in the mean of order two.*

The sequence $f(x)$ is said to be *convergent in the mean of order* $p$ $(p > 0)$ to a function $f(x)$ if

$$\int_E |f_n(x) - f(x)|^p \, dx \to 0 \quad \text{as} \quad n \to \infty.$$

We most frequently deal with the case $p \geqslant 1$.

The following theorem holds.

LEBESGUE'S THEOREM. *Suppose that on a measurable set $E$ we are given a sequence $\{f_n(x)\}$ of measurable and almost everywhere finite functions, which converges almost everywhere on $E$ to a function $f(x)$. Then for any $\varepsilon > 0$*

$$\lim_{n \to \infty} mE\{|f_n(x) - f(x)| \geqslant \varepsilon\} = 0.$$

This theorem leads to the introduction of the following definition. Suppose we are given a sequence of functions $\{f_n(x)\}$ measurable and almost everywhere finite on a measurable set $E$. We say that it *converges in measure on the set $E$ to the function $f(x)$ if*

$$\lim_{n \to \infty} mE\{|f_n(x) - f(x)| \geqslant \varepsilon\} = 0,$$

for any $\varepsilon > 0$. Then Lebesgue's theorem cited above may be formulated as follows: a sequence, *which converges almost everywhere, converges in measure to the same limit function.* It is obvious that we may also consider any function, which coincides with $f(x)$ almost everywhere, to be *a limit function, in the sense of convergence in measure* for a given sequence $\{f_n(x)\}$.

Lebesgue's theorem cannot be reversed, i.e. from the convergence of a sequence in measure its convergence almost everywhere does not necessarily follow.

Moreover, *a sequence may converge in measure although it has no point of convergence.* To be convinced of the existence of such a sequence, it is sufficient to consider the following example. We shall define, for each natural number $k$, a group of functions, $f_1^k(x)$, $f_2^k(x)$ ..., $f_k^{(k)}(x)$, assuming that

$$f_i^{(k)}(x) = \begin{cases} 1 & \text{if } x \in \left[\dfrac{i-1}{k}, \dfrac{i}{k}\right], \\ 0 & \text{outside this segment.} \end{cases}$$

Having numbered in succession all the functions we have constructed, we obtain a sequence $\{\varphi_n(x)\}$, which *converges in measure* to the function which is identically equal to zero on [0, 1].

Now let us suppose that $\varphi_n(x) = f_i^{(k)}(x)$. It is sufficient to consider the case $0 < \varepsilon \leqslant 1$. Then

$$E\{|\varphi_n(x)| \geqslant \varepsilon\} = \left[\frac{i-1}{k}, \frac{i}{k}\right],$$

so that $mE\{|\varphi_n(x)| \geqslant \varepsilon\} = \frac{1}{k} \to 0$ as $n \to \infty$. Also this sequence does not converge at any one point, because for any $x_0 \in [0, 1]$ and any $k > 2$ we can find $i_1$ and $i_2$ such that

$$x_0 \in \left[\frac{i_1-1}{k}, \frac{i_1}{k}\right], \qquad x_0 \overline{\in} \left[\frac{i_2-1}{k}, \frac{i_2}{k}\right],$$

so that $f_{i_1}^{(k)}(x_0) = 1$, $f_{i_2}^{(k)}(x_0) = 0$. Thus at each point $x_0$ of the function, the sequences $\{\varphi_n(x)\}$ take on both the value 1 and the value 0 an infinite number of times, so that the sequence $\{\varphi_n(x)\}$ *diverges at each point of* [0, 1].

It is possible to draw a conclusion from convergence in measure about convergence almost everywhere, but in the following strict form: *if a sequence* $\{f_n(x)\}$ *converges in measure to a function* $f(x)$ *on a set* $E$, *then we may choose a subsequence*

$$f_{n_1}(x), \quad f_{n_2}(x), \quad f_{n_3}(x), \ldots \quad (n_1 < n_2 < n_3 < \cdots)$$

*from it, which converges to* $f(x)$ *almost everywhere on* $E$.

*A sequence of functions, which converges in the mean of order* $p$ ($p > 0$) *converges in measure to the same function.* Therefore, *for the sequence which converges in the mean, there also exists a sequence which is convergent almost everywhere.* As we have seen above, however, *the sequence itself, which converges in the mean, may turn out to be divergent at each point.* In exactly the same way *convergence in the mean does not necessarily follow from convergence in measure, since convergence at each point does not imply convergence in the mean.*

This is confirmed by the sequence $\{f_n(x)\}$, defined on [0, 1] by the equations

$$f_n(x) = \begin{cases} n, & \text{if} \quad 0 < x < \dfrac{1}{n}, \\ 0 & \text{at other points of } [0, 1]. \end{cases}$$

It is obvious that $\lim_{n\to\infty} f_n(x) = 0$ for any $x \in [0, 1]$, so that the sequence $\{f_n(x)\}$ converges to zero at each point, but

$$\int_0^1 f_n^2(x)\,dx = \int_0^{\frac{1}{n}} n^2\,dx = n \to \infty,$$

and consequently convergence in the mean does not hold.

**5.** Let us consider now the conditions for which passage to the limit under the integral sign is possible in a Lebesgue integral. As in the Riemann integral (see § 6, para. 2), to permit passage to the limit under the integral sign convergence at each point is not a sufficient condition, even in the case where the limit function is integrable. We can easily see this by considering the previous example. Actually with any $x \in [0, 1]$ we have $\lim_{n\to\infty} f_n(x) = 0$. On the other hand,

$$\int_0^1 f_n(x)\,dx = 1,$$

and this integral does not approach zero.

For bounded functions we have the following theorem.

*Suppose that on a measurable set $E$ a sequence $\{f_n(x)\}$ of bounded measurable functions is given, which converges on $E$ in measure to a measurable bounded function $f(x)$. If the sequence $\{f_n(x)\}$ is uniformly bounded, i.e. there exists a constant $M$, such that for all $n$ and for all $x$, $|f_n(x)| < M$, then*

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx.$$

The following statement puts it more forcibly.

*Let a sequence $\{f_n(x)\}$ of summable functions converge in measure on a set $E$ to a summable function $f(x)$. If all functions of the sequence are dominated by a summable function, i.e. there exists a summable function $\varphi(x)$ such that, for all $n$ and all $x \in E$,*

$$|f_n(x)| \leqslant \varphi(x),$$

*then*

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx.$$

As we have already remarked, a sequence which converges almost everywhere is convergent in measure. Therefore both the

theorems formulated above remain in force if, in them, we replace convergence in measure by convergence almost everywhere.

## § 7. ORTHOGONAL SYSTEMS OF FUNCTIONS

**1.** We shall limit ourselves to the consideration only of measurable functions $f(x)$, defined on a segment $[a, b]$, whose square is summable on this segment, i.e. the Lebesgue integral $\int_a^b [f(x)]^2\,dx$ exists and is finite. The set of all such functions is usually denoted by $L^2_{[a,\,b]}$, or simply $L^2$.†

The system of functions $\{\varphi_n(x)\}$, belonging to $L^2$, is said to be *orthogonal* on the segment $[a, b]$ if

$$\int_a^b \varphi_m(x)\,\varphi_n(x)\,dx = 0 \quad (m \neq n).$$

We usually consider *normalized* systems whose functions possess the property

$$\int_a^b [\varphi_n(x)]^2\,dx = 1.$$

If the system $\{\varphi_n(x)\}$ is orthogonal and normalized then it is called, for short, an *orthonormal system*. Thus an orthonormal system is defined by the relations

$$\int_a^b \varphi_m(x)\,\varphi_n(x)\,dx = \delta_{m,n} = \begin{cases} 0 & \text{if} \quad m \neq n, \\ 1 & \text{if} \quad m = n. \end{cases}$$

The symbol $\delta_{m,n}$ is called *Kronecker's symbol.*

*An orthonormal system of functions in $L^2$ is at most denumerable.*

A system of functions $\{\varphi_n(x)\}$ is said to be *complete* with respect to $L^2$, if from the relations $f \in L^2$ and

$$\int_a^b \varphi_n(x)\,f(x)\,dx = 0 \quad (n = 1, 2, \cdots)$$

it follows that $f(x) = 0$ almost everywhere, i.e. if in $L^2$ there exists no function which is orthogonal to all functions of the given system, except those which are equal to zero almost everywhere.

† For more detail, see § 9.

The system of functions $\{\varphi_n(x)\}$ is *closed* with respect to $L^2$, if any function $f(x)$ with a summable square may be approximated in the mean by a linear combination of the system, with any degree of accuracy. To be more precise: for each function $f(x) \in L^2$ and any $\varepsilon > 0$ we can find $n$ functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ of our system and also real numbers $c_1, c_2, \ldots, c_n$, such that

$$\int_a^b [f(x) - c_1\varphi_1(x) - \cdots - c_n\varphi_n(x)]^2\, dx < \varepsilon^2.$$

The ideas of completeness and closure in $L^2$ are equivalent, i.e. any closed system is complete and any complete system is closed. A complete or closed system is always infinite, therefore a *complete* (closed) *orthonormal system is always infinite, but denumerable.*

If a system is complete or closed on a segment $[a, b]$, then it retains this property also for any inner segment. *Any system which is orthonormal on* $[a, b]$ *may be extended to a complete orthonormal system by adding suitable functions.*

Let $\{\varphi_n(x)\}$ be an orthonormal system of functions. The series

$$\sum_{n=1}^{\infty} a_n\varphi_n(x),$$

where $a_n$ is any constant, is called an *orthogonal series.*

If this series converges uniformly on the segment $[a, b]$ to the function $f(x)$, then the coefficients $a_n$ of the series are easily determined directly. Multiplying both sides of the equation

$$f(x) = \sum_{n=1}^{\infty} a_n\varphi_n(x)$$

by $\varphi_k(x)$ and then integrating on the segment $[a, b]$, we find that

$$\int_a^b f(x)\,\varphi_k(x)\,dx = \sum_{n=1}^{\infty} a_n \int_a^b \varphi_n(x)\,\varphi_k(x)\,dx = a_k.$$

The coefficients, obtained in this formula, are called *Fourier coefficients* of the function $f(x)$ with respect to the orthonormal system $\{\varphi_n(x)\}$. If all the coefficients of an orthognal series are Fourier coefficients of the function $f(x)$, the series is said to be a *Fourier expansion* or *Fourier series* of the function $f(x)$ with respect to the system $\{\varphi_n(x)\}$. It has been shown above that *if an orthogonal series converges uniformly to the function* $f(x)$, *then it is*

*a Fourier series of this function.* It is not difficult to prove the correctness of this statement also for the case of *convergence in the mean.*

The definition of a Fourier series essentially depends on the fact that in some sense integration is performed. In this connection series are considered as Fourier–Riemann, Fourier–Stieltjes or Fourier–Lebesgue, etc.

If the system $\{\varphi_n(x)\}$ is not normalized, the expression for the Fourier coefficients is slightly altered. For instance, if

$$\int_a^b [\varphi_n(x)]^2\,dx = \lambda_n > 0,$$

the system

$$\{\psi_n(x)\} = \left\{\frac{1}{\sqrt{\lambda_n}}\,\varphi_n(x)\right\}$$

is already normalized. Let $b_k$ represent the $k$th coefficient of the function $f(x)$ with respect to the system $\{\psi_n(x)\}$. Then

$$b_k = \int_a^b f(x)\,\psi_k(x)\,dx = \int_a^b f(x)\,\frac{\varphi_k(x)}{\sqrt{\lambda_k}}\,dx = \frac{a_k}{\sqrt{\lambda_k}},$$

where we assume as before that

$$a_k = \int_a^b f(x)\,\varphi_k(x)\,dx.$$

The expansion of the function $f(x)$ with respect to the system $\{\psi_n(x)\}$ gives

$$\sum_{k=1}^{\infty} b_k\psi_k(x) \equiv \sum_{k=1}^{\infty} \frac{a_k}{\sqrt{\lambda_k}}\,\frac{\varphi_k(x)}{\sqrt{\lambda_k}} \equiv \sum_{k=1}^{\infty} \frac{a_k\varphi_k(x)}{\lambda_k},$$

from which it is obvious that the Fourier coefficients with respect to the system $\{\varphi_n(x)\}$, which is not normalized, are equal to

$$\frac{1}{\lambda_k}\int_a^b f(x)\varphi_k(x)\,dx.$$

This formula may also be obtained by direct integration.

2. Fourier coefficients may be obtained as a solution of the problem of *the best approximation in the mean.* Suppose that on $[a, b]$

a function $f(x) \in L^2$ is defined and $\{\varphi_n(x)\}$ is an orthonormal system of functions from $L^2$. Let us assume that

$$s_n(x) = \sum_{k=1}^{n} \xi_k \varphi_k(x)$$

and let us pose the problem of selecting, with constant $n$, coefficients $\xi_k$ such that the integral

$$\int_a^b [f(x) - s_n(x)]^2 \, dx$$

is a minimum. *A solution of this problem is a partial sum of a Fourier series*, i.e. the approximation in the mean will be best when the coefficients $\xi_k$ are equal to the respective Fourier coefficients.

For any function $f(x) \in L^2$ *Bessel's inequality* is true, i.e.

$$\sum_{k=1}^{\infty} a_k^2 \leqslant \int_a^b [f(x)]^2 \, dx,$$

where $\{a_k\}$ denotes the Fourier coefficients of the function $f(x)$ with respect to the system $\{\varphi_n(x)\}$. Hence it follows that *for a function with a summable square the series composed of the squares of the Fourier coefficients converges.* In particular, *the Fourier coefficients approach zero.*

*The orthogonal series* $\sum_{k=1}^{\infty} a_k \varphi_k(x)$ *converges in the mean when, and only when,* $\sum_{k=1}^{\infty} a_k^2 < \infty$.

Let $\{\varphi_n(x)\}$ be an orthonormal system, complete with respect to $L^2$ and $f(x) \in L^2$. Then *the Fourier series of the function $f(x)$ with respect to the system $\{\varphi_n(x)\}$ converges in the mean to the function $f(x)$.* For a complete system *Parseval's equation* is also valid, viz.

$$\sum_{k=1}^{\infty} a_k^2 = \int_a^b [f(x)]^2 \, dx.$$

THE RIESZ–FISCHER THEOREM. *If a system $\{\varphi_n(x)\}$ is orthonormal, and $\{a_i\}$ is a sequence of numbers with the sum of its squares convergent, then there exists a function $f(x) \in L^2$ for which the series* $\sum_{k=1}^{\infty} a_k \varphi_k(x)$ *is a Fourier series with respect to the system $\{\varphi_n(x)\}$.*

A similar function is determined in unique form to within its value on a set of measure zero when, and only when, $\{\varphi_n(x)\}$ is complete.

The following theorem gives a sufficient condition for the convergence almost everywhere of an orthogonal series in the general case.

THE THEOREM OF D. E. MEN'SHOV AND G. RADEMACHER. *Let* $\{\varphi_n(x)\}$ *be an orthonormal system from* $L^2$ *on* $[a, b]$ *and* $\sum_{n=1}^{\infty} a_n^2 \ln^2 n < \infty$. *Then the orthogonal series* $\sum_{n=1}^{\infty} a_n \varphi_n(x)$ *converges almost everywhere on* $[a, b]$.

For some orthogonal systems more forcible statements are correct. For instance, we have the following theorems:

THE THEOREM OF A. N. KOLMOGOROV, A. SELIVERSTOV AND A. I. PLESSNER. *If the series* $\sum_{n=1}^{\infty} (a_n^2 + b_n^2) \ln n$ *converges, then the trigonometric series* $\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$ *converges almost everywhere.*

G. RADEMACHER'S THEOREM. *If* $\sum_{n=1}^{\infty} a_n^2 < \infty$, *then the series* $\sum_{n=1}^{\infty} a_n r_n(x)$, *where* $r_i(x)$ *is the Rademacher function* (see para. 3 below) *converges almost everywhere on the segment* $[0, 1]$.

A. N. KOLMOGOROV'S THEOREM. *If* $\sum_{n=1}^{\infty} a_n^2 = \infty$, *then the series* $\sum_{n=1}^{\infty} a_n r_n(x)$ *diverges almost everywhere on* $[0, 1]$.

**3.** The most frequently met example of an orthogonal system is the *trigonometrical system*

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x, \ldots,$$

which is a *complete* (*closed*) *orthogonal system* on the segment $[0, 2\pi]$ or $[-\pi, \pi]$, and also on any other segment of the form $[\alpha, \alpha + 2\pi]$. This system is not normalized, since $\int_0^{2\pi} 1^2 dx = 2\pi$ and $\int_0^{2\pi} \cos^2 nx dx = \int_0^{2\pi} \sin^2 nx dx = \pi$. Therefore the *normalized trigonometric system* has the form

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos x, \frac{1}{\sqrt{\pi}} \sin x, \frac{1}{\sqrt{\pi}} \cos 2x, \frac{1}{\sqrt{\pi}} \sin 2x, \ldots.$$

The system

$$1, \quad \cos\frac{\pi x}{l}, \quad \sin\frac{\pi x}{l}, \quad \cos\frac{2\pi x}{l}, \quad \sin\frac{2\pi x}{l}, \ldots,$$

which is orthogonal on $[-l, l]$ or on any segment of length $2l$, is also called a trigonometrical system, as are also the systems

$$\{\cos nx\} \quad (n = 0, 1, 2, \ldots) \quad \text{and} \quad \{\sin nx\} \quad (n = 1, 2, \ldots).$$

These latter are *complete* orthogonal systems on the segment $[0, \pi]$. It is easy to see that they are orthogonal also on $[0, 2\pi]$, but here they cease to be complete.

Another example of an orthogonal system on the segment $[-1, 1]$ is the complete system of Legendre polynomials, $\{P_n(x)\}$, where

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2 - 1)^n.$$

To normalize this function, it is sufficient to multiply the polynomial $P_n(x)$ by $\sqrt{\dfrac{2n+1}{2}}$.

*Rademacher functions*, which form an orthonormal system on the segment $[0, 1]$, may be formed in the following way. Let $r_0(x) \equiv 1$. For a natural number $k > 0$ we divide the segment $[0, 1]$ into $2^k$ equal segments, on each of which the function $r_k(x)$ takes alternately the values $+1$ and $-1$ and at their ends the value zero. We may also say that for the point $x \in [0, 1]$ the function $r_k(x)$ takes the value $+1$ if the $k$th place of the binary expansion of $x$ contains zero, and the value $-1$ if the $k$th place of the binary expansion of $x$ contains unity; if $x$ has two binary expansions with different digits in the $k$th place, then $r_k(x) = 0$.

This system possesses the interesting property of *multiplicativity: if $j \leqslant k \leqslant l \leqslant \cdots \leqslant q$ are natural numbers and $r_i(x)$ Rademacher functions, then*

$$\int_0^1 r_j(x)\, r_k(x) \cdots r_q(x)\, dx = 0,$$

*except in the case where the expression under the integral sign is the product of pairs of identical factors. In this case the integral is equal to unity.*

Rademacher functions admit of probability interpretation. The function $r_1(x)$ is equal to $+1$ for $x \in (0, \frac{1}{2})$ and $-1$ for $x \in (\frac{1}{2}, 1)$. If we choose at random a point $x$, considering the probability of

the choice of a point belonging to the set $E$, which is $mE$, the probabilities of the values $+1$ and $-1$ for the function $r_1(x)$ are identical and equal to $\frac{1}{2}$, which corresponds to the probabilities of the occurrence of heads and tails with one toss of a coin. Similarly the problem of $n$ tosses, or any other problem on repeated trials with two equally probable results, may be formulated with the help of the functions $r_n(x)$. In this connection all theorems, which are related to Rademacher functions, may be interpreted in the theoretic-probability sense.

Rademacher's system is an incomplete system. Its natural complement is the *system of Walsh functions* $\{w_n(x)\}$, which may be formed as follows. Let a number $n$ be given in the form of a sum of powers of two, $n = \sum_{n=0}^{m} a_k 2^k$, where $a_k$ takes the values 0 and 1. Let us assume that $w_0(x) \equiv r_0(x) \equiv 1$ and

$$w_n(x) = \prod_{k=0}^{m} [r_{k+1}(x)]^{a_k} \quad n \geqslant 1,$$

where $r_i(x)$ is the corresponding Rademacher function. In particular

$$w_1(x) = r_1(x); \quad w_2(x) = r_2(x); \quad w_3(x) = r_1(x)\, r_2(x); \ldots$$

$$w_7(x) = r_1(x)\, r_2(x)\, r_3(x); \ldots$$

The system of Walsh functions is a *complete orthonormal system* on the segment [0, 1].

4. Let the system of functions $\{\varphi_n(x)\}$ with some fixed nonnegative function $w(x)$ posses the property

$$\int_a^b w(x)\, \varphi_m(x)\, \varphi_n(x)\, dx = \delta_{m,n}.$$

Such a system is usually called *orthogonal with weight* $w(x)$. It is clear that with this system $\{\varphi_n(x)\sqrt{w(x)}\}$ will be orthogonal in the usual sense. The function $w(x)$ is called a *weight function*. With $w(x) \equiv 1$ we have the idea of orthogonality discussed above.

Systems of functions which are orthogonal with a weight often arise in the consideration of boundary value problems of ordinary linear differential equations of the second order. Let us consider

an equation of the second order of the form

$$\frac{d^2y}{dx^2} + \lambda A(x)\, y = 0,$$

where $\lambda$ is a real parameter and $A(x) > 0$ is a continuous function. *Characteristic values* of an equation are those values of the parameter $\lambda$, for which solutions exist, which are not identically zero and which satisfy the boundary conditions $y(a) = y(b) = 0$. Functions which satisfy an equation with characteristic values of the parameter are called *characteristic functions*. If $y_i(x)$, $y_k(x)$ are two characteristic functions, then

$$\int_a^b A(x)\, y_i(x)\, y_k(x)\, dx = 0.$$

Thus the *characteristic functions form a system of functions which are orthogonal with a weight* $A(x)$.

Among the functions, which are orthogonal with a weight, an important part is played by polynomials. *For each weight function* $w(x)$, *which is summable on the segment* $[a, b]$ *and positive almost everywhere on it, there exists a system of polynomials which are orthogonal with weight* $w(x)$ *on* $[a, b]$, *and uniquely defined except for factors* $\pm 1$. For systems of functions, which are orthogonal with a weight $w(x)$, the basic facts hold, as established in § 7, paras. 1 and 2. Here instead of functions from $L^2$ we must consider functions $f(x)$ for which

$$\int_a^b f^2(x)\, w(x)\, dx < \infty,$$

and the Fourier coefficients of the function $f(x)$ with respect to the system $\{\varphi_n(x)\}$, which is orthogonal with a weight $w(x)$, are defined by the equation

$$a_n = \int_a^b w(x)\, f(x)\, \varphi_n(x)\, dx.$$

More extensively used systems of polynomials, which are orthogonal with a weight, are considered in the publication in the Mathematical Reference Library *Mathematical analysis* (*functions, limits, series, continued fractions*).

**5.** Let $f_1(x), f_2(x), \ldots, f_n(x)$ be a system of $n$ functions from $L^2$, which are defined on $[a, b]$. This system is said to be *linearly independent* on this segment, if no linear combination of these functions, with coefficients different from zero, is equal to zero almost everywhere on $[a, b]$.

The criterion for the linear independence of the functions $f_1(x), f_2(x), \ldots, f_n(x)$ is given by the following theorem.

GRAM'S THEOREM. *In order that the functions $f_1(x), f_2(x), \ldots, f_n(x)$ may be linearly independent, it is necessary and sufficient that the determinant of the matrix $\|a_{ik}\|$, whose elements are defined by the formulae*

$$a_{ik} = \int_a^b f_i(x) f_k(x)\, dx \quad (i, k = 1, 2, \ldots, n)$$

*(the Gram determinant) should be different from zero.*

We shall denote the Gram determinant of the system $f_1(x), f_2(x), \ldots, f_n(x)$ by $G(f_1, f_2, \ldots, f_n)$. Two systems of functions $f_1(x), \ldots, f_n(x)$ and $g_1(x), \ldots, g_n(x)$ are said to be *equivalent* if any of the functions of one system can be given in the form of a linear combination of functions of the other system. The problem of the *orthogonalization* of a system can be put in the following form; given a linearly independent system $f_1(x), \ldots, f_n(x)$, is it possible to find an equivalent system $\varphi_1(x), \ldots, \varphi_n(x)$, consisting of orthogonal functions? This problem always has a solution. We may carry out the orthogonalization of the system $f_1(x), \ldots, f_n(x)$ in the following way. Let us suppose that

$$\begin{aligned}
\varphi_1(x) &= c_{11} f_1(x), \\
\varphi_2(x) &= c_{21} f_1(x) + c_{22} f_2(x), \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\varphi_n(x) &= c_{n1} f_1(x) + c_{n2} f_2(x) + \cdots + c_{nn} f_n(x),
\end{aligned}$$

such that $c_{ik} = 0$ with $k > i$, and that we successively choose coefficients $c_{ik}$ with $k \leqslant i$ so that the functions $\varphi_1(x), \ldots, \varphi_n(x)$ may be orthogonal and normalized. The functions $\varphi_k(x)$ are expressed by the formula

$$\varphi_k(x) = \frac{1}{\sqrt{G_k G_{k-1}}} \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ a_{k-1,1} & a_{k-1,2} & \cdots & a_{k-1,k} \\ f_1(x) & f_2(x) & \cdots & f_k(x) \end{vmatrix},$$

where $G_k = G(f_1, f_2, \ldots, f_k)$ $(k \geqslant 1)$ and $G_0 \equiv 1$.

The above process is called *E. Schmidt's process of orthogonalization.*

In a number of cases it is interesting to put the problem of orthogonalization in another way. For instance, if we have a system of linearly independent functions $\{f_n(x)\}$, given on a segment $[a, b]$, we are required to find an extension of these functions on a larger segment, so that the extended system proves to be orthogonal.

*If $\{f_n(x)\}$ is an arbitrary system of functions from $L^2$, defined on a segment $[a, b]$, the function $f_n(x)$ may be so extended on the segment $[b, c]$ that the system obtained is orthogonal on $[a, c]$.* If the original system is orthogonal on $[a, b]$, then after extension it will also be orthogonal on $[b, c]$.

The system of functions obtained as a result of orthogonalization may turn out not to be normalized, and its normalization by multiplication by constants may change the function $f_n(x)$ on $[a, b]$. Therefore the question of the possibility of extending the system to an orthonormal one is natural. The conditions required to make such an extension possible are given by the following theorem.

I. Schur's Theorem. *Let functions $\{f_n(x)\}$ be defined on $[a, b]$, $0 < a < b < 1$ and $f_n(x) \in L^2$. In order that there may exist on $[0, 1]$ an orthonormal system $[\varphi_n(x)]$, for which $\varphi_n(x) \equiv f_n(x)$ with $x \in [a, b]$, it is necessary and sufficient that the maximum of the integral*

$$\int_a^b \left[\sum_{i=1}^{\infty} \xi_i f_i(x)\right]^2 dx$$

*with any different sequences $\{\xi_i\}$, satisfying the condition $\sum_{i=1}^{\infty} \xi_i^2 = 1$, should not exceed unity.*

6. In some applied questions we have to consider systems which are orthogonal, or orthogonal with a weight, on infinite intervals $(0, \infty)$ or $(-\infty, \infty)$. The properties of these systems are similar to the properties of the systems given on a finite interval. Orthogonality is sometimes defined differently. For instance, the system of functions $\{\varphi_n(x)\}$ is said to be *orthogonal* on $(-\infty, \infty)$ if

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \varphi_m(x)\varphi_n(x)\, dx = 0 \quad (T > 0,\ m \neq n).$$

It is easy to see that for orthogonality in this sense, uniform boundedness of the integrals of products of the functions is sufficient. In particular, systems which consist of sines and cosines possess this property. Thus the system $\{\cos \alpha x\}$, $(-\infty < \alpha < \infty)$, is orthogonal in this sense on $(-\infty, \infty)$ so that in this case an orthogonal system may prove to be non-denumerable.

## § 8. FUNCTIONS OF SEVERAL VARIABLES

**1.** The basic ideas of § 1 may without difficulty be transferred to plane point sets or point sets in a space of any number of dimensions. The definitions of operations on these sets and the notation for these operations hold in the same way.

The definitions of isolated and limit points of a set of closed, perfect and open sets are also valid, but the term "neighbourhood" takes on another meaning. If, for a linear set, we understand by the neighbourhood of a point any interval which contains it, by *the $\varepsilon$-neighbourhood of a point of a plane set* we mean an open circle of radius $\varepsilon$ with centre at the given point.

To conveniently transfer this definition to a space of any number of dimensions we shall give an algebraic definition: the *$\varepsilon$-neighbourhood of a point $M(a, b)$ on a plane* is the set of points $\{(x, y)\}$, which satisfy the inequality $(x-a)^2 + (y-b)^2 < \varepsilon^2$. Similarly *the $\varepsilon$-neighbourhood of a point $M(a, b, c)$ in space* is an open sphere of radius $\varepsilon$ with centre at the given point, i.e. the set of points $\{(x, y, z)\}$ which satisfy the inequality $(x-a)^2 + (y-b)^2 + (z-c)^2 < \varepsilon^2$.

As is well known, see the Mathematical Reference Library *Mathematical analysis (functions, limits, series, continued fractions)*, *a point of space of n dimensions* is an ordered system of $n$ real numbers $(x_1, x_2, \ldots, x_n)$, called its *coordinates*. The set of all possible $n$-dimensional points will be called *Euclidean space of n dimensions*, if the distance between the points $x(x_1, \ldots, x_n)$ and $y(y_1, \ldots, y_n)$ is defined by the formula $\varrho(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$. Euclidean $n$-dimensional space is usually denoted by the letters $E^n$ or $R^n$.

From now on, unless it is clearly stipulated otherwise, we shall consider point sets in Euclidean space $R^n$. The part of segments is played in space $R^n$ by parallelepipeds, i.e. sets of points $\{M\}$, $M \in R^n$ whose coordinates satisfy the inequalities $a_k \leqslant x_k \leqslant b_k$ $(k = 1, 2, \ldots, n)$.

*A closed* (*open*) *n-dimensional sphere* of radius $r$ and centre at the point $a$ is defined as the set of points $x$ which satisfy the conditions $\varrho(x, a) \leqslant r$ ($\varrho(x, a) < r$). The open sphere of radius $\varepsilon$ and centre at the point $a$ is the corresponding $\varepsilon$-neighbourhood of the point $a$.

As distinct from linear sets the *domain* in $n$-dimensional space ($n \geqslant 2$) is not any open set, but an open and connected set. An *open* set is said to be *connected* it any two of its points can be joined by a polygonal line, all of whose points belong to this set.

For an open space set we cannot define constituent intervals as for linear sets. Only the following statement is correct: *any non-empty open set is the sum of a denumerable set of closed parallelepipeds, which have no common inner points.*

**2.** A variable quantity $y$ is said to be a function of point of the space $R^n$ or a *function of n variables* $y = f(M) = f(x_1, x_2, \ldots, x_n)$, defined on a set $E \subset R^n$, if there is a definite value of y corresponding to each point $M(\equiv(x_1, x_2, \ldots, x_n)) \in E$.

Let $\sigma$ be some system of sets of the space $R^n$. We say that *a function of the set is defined on* $\sigma$ if a definite number $\Phi(E)$ corresponds to each $E \in \sigma$.

A function of sets $\Phi(E)$ is *additive* on $\sigma$, if, for any two disjoint sets $E_1, E_2 \in \sigma$, $\Phi(E_1 + E_2) = \Phi(E_1) + \Phi(E_2)$. It is obvious that if $\Phi(E)$ is additive, then for any number of mutually disjoint sets $E_1, E_2, \ldots, E_m$

$$\Phi\left(\sum_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{m} \Phi(E_k).$$

If this property is also true for a denumerable set of mutually disjoint sets $\{E_k\}$, i.e.

$$\Phi\left(\sum_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \Phi(E_k)$$

the function of the sets $\Phi(E)$ is said to be *completely additive.*

The simplest completely additive function of sets is *volume* (more accurately *n-dimensional volume*; for $n = 2$ it is *area*). For an $n$-dimensional parallelepiped $S \subset R^n$, the coordinate of whose points $M \in S$ satisfy the inequality $a_k \leqslant x_k \leqslant b_k$ ($k = 1, 2, \ldots, n$) where $M \equiv (x_1, x_2, \ldots, x_n)$, *the volume is defined by the equation*

$$v(S) = \prod_{k=1}^{n} (b_k - a_k).$$

This definition allows us to ascribe volume to any set which can be obtained from a denumerable set of parallelepipeds by means of the operations of addition, intersection and taking the complement.

A commoner example of a completely additive function of a set is *measure*. The measure of a space set for a system of sets $\sigma$ is defined as *a non-negative, completely additive and monotonic function of the sets*, i.e. a function $\mu(E)$, for which

(1) $\mu(E) \geqslant 0$ for any $E \in \sigma$,

(2) $\mu(E_1) \leqslant \mu(E_2)$ if $E_1 \subseteq E_2 (E_1, E_2 \in \sigma)$,

(3) $\mu\left(\sum_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$ if $E_1, E_2, \ldots$ are mutually disjoint sets from $\sigma$.

A more commonly used measure for space sets is the Lebesgue measure, which may be obtained by means of the following construction. Let $E$ be a bounded point set in $R^n$, lying inside a parallelepiped $S$, $E \subseteq S$ and $G$ an open set in $S$, which contains $E$, $S \supseteq G \supseteq E$. The value of the volume $v(G)$ in the form of the sum of a denumerable set of closed parallelepipeds, is defined uniquely and does not depend on the method of choice of the parallelepipeds. The lower bound, $\inf v(G)$, for all sets $G$ of $S$, which contain $E$, will be called the *outer n-dimensional Lebesgue measure* of the set $E$ and will be written

$$m_e E = \inf v(G).$$

The set $E$ is said to be *measurable* if

$$m_e E + m_e CE = v(S),$$

where $CE$ is the complement of the set $E$ with respect to the parallelepiped $S$, $CE = S - E$. The value of the outer measure is also called *the n-dimensional Lebesgue measure of the measurable set* $E$. The quantity

$$m_i E = v(S) - m_e CE$$

is called *the inner n-dimensional Lebesgue measure* of the set $E$. It may also be defined as the upper bound of the volume of closed sets $F$, contained in $E$,

$$m_i E = \sup v(F), \quad F \subseteq E.$$

Let $E$ be a plane set, consisting of points $(x, y)$. By the *section* of the set $E$ by the straight line $x = x_0$ we mean a linear set $E(x_0)$, which consists of numbers $y$, such that the point $(x_0, y) \in E$. The plane measure of $E$ is connected with the linear measure of its section and the following theorem applies.

*Let $E$ be a plane set which is contained in an open rectangle $S(a < x < b, c < y < d)$. Then*

(a) *for almost all $x \in (a, b)$ the set $E(x)$ is measurable;*

(b) *if $\Delta$ denotes the set of those $x \in (a, b)$, for which $E(x)$ is measurable ($m\Delta = b - a$), the function $mE(x)$ is measurable on $\Delta$;*

(c) $mE = \int_{\Delta} mE(x)\,dx$,

*where the integral must be understood in the Lebesgue sense.*

From this theorem it follows that, *if a plane set $E$ is of measure zero, then almost all its sections are sets of measure zero.* Conversely, *if almost all sections of a plane, measurable set have measure zero, then $mE = 0$.*

This theorem can also be applied to the case of a set in space $R^n$.

Besides the Lebesgue measure for space sets we sometimes use the *Hausdorff k-dimensional measure*, which is defined in the following way. Let $E$ be a set in $n$-dimensional space $R^n$, $\sigma$ a finite system of $n$-dimensional spheres enclosing $E$, whose diameters do not exceed $\delta$, and $\Sigma_\delta d$ the sum of the diameters of the spheres, which belong to the system $\sigma$.

Then the limit,

$$\lim_{\delta \to 0} \inf \Sigma_\delta d = \nu_1(E),$$

where the lower limit is taken over all possible systems $\sigma$, always exists, although it may turn out to be infinite.

The quantity $\nu_1(E)$ is called the *Hausdorff linear measure* or the *Hausdorff length* of the $n$-dimensional set $E$. To obtain the *Hausdorff plane (two-dimensional) measure*, instead of $\Sigma_\delta d$ we must take $\Sigma_\delta \frac{\pi d^2}{4}$, i.e. the sum of the areas of two dimensional great circles. Similarly, we may define the Hausdorff measure $\nu_k(E)$ of any degree $k$ for $1 \leqslant k \leqslant n$. It is clear that if $\nu_k(E)$ is finite and distinct from zero, then $\nu_m(E) = 0$ if $m > k$ and $\nu_m(E) = \infty$ if $m < k$.

**3.** Let $M \equiv (x_1, x_2, \ldots, x_n)$ denote a point of space $R^n$ and $y = f(M)$ a function of $n$ variables, defined on a set $E \subset R^n$. We

say that the function $f(M)$ is continuous at the point $M_0 \in E$ *with respect to the set* $E$, if for any sequence of points $\{M_m\}$ $(M_m = (x_1^{(m)}, x_2^{(m)}, \ldots, x_2^{(m)}) \in E)$ converging to $M_0$, i.e. such that $\varrho(M_m, M_0) \to 0$,

$$\lim_{m \to \infty} f(M_m) = f(M_0).$$

If the function $f(M)$ is given and continuous on a closed set $F$, the sets $E\{f \geqslant a\}$ and $E\{f \leqslant a\}$ are closed for any $a$.

The function $f(M)$ is said to be *measurable* on a set $E$ if $E$ is Lebesgue measurable and the sets $E\{f < a\}$ are measurable for any $a$. In particular, *a function, which is given and continuous on a closed parallelepiped, is measurable.* For a measurable function of $n$ variables we have the theorems of E. Borel and N. N. Luzin on the $C$-property (see § 3, para. 5).

A function of two variables $f(x, y)$ is said to be *differentiable* at the point $(x_0, y_0)$, if there exist numbers $A$ and $B$ such that the relation

$$\frac{f(x, y) - f(x_0, y_0) - A(x - x_0) - B(y - y_0)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}}$$

approaches zero as $(x, y) \to (x_0, y_0)$. The expression $A(x - x_0) + B(y - y_0)$ or the number pair $\{A, B\}$ is said to be the *exact total differential* of the function $f$ at the point $(x_0, y_0)$. In this case the numbers $A$ and $B$ are the *partial derivatives* of the function $f(x, y)$ at the point $(x_0, y_0)$ with respect to $x$ and $y$.

If this ratio approaches the limit *asymptotically*† for some $A$ and $B$, the number pair $\{A, B\}$ is called the *asymptotic differential* of the function $f(x, y)$ at the point $(x_0, y_0)$, and the numbers $A$ and $B$ are the asymptotic differential *coefficients*.

V. V. Stepanov's Theorem. *In order that a finite function of two variables $f(x, y)$, which is measurable on a set $E$, may be asymptotically differentiable at almost all points of $E$, it is necessary and sufficient that it should be asymptotically differentiable with respect to each of the variables almost everywhere on $E$.*

*If $f(x, y)$ is asymptotically differentiable almost everywhere on $E$, then the coefficients of the asymptotic differential coincide with the asymptotic partial derivatives $f_x^{[1]}(x, y)$, $f_y^{[1]}(x, y)$ at almost every point of $E$.*

† i.e. there exists a plane set of positive measure, which has $(x_0, y_0)$ as a density point with respect to the two-dimensional Lebesgue measure, for which there exists a limit in the ordinary sense.

**4.** Let a function of two variables $f(x, y)$ be defined and continuous on the rectangle $S(a \leqslant x \leqslant b,\ c \leqslant y \leqslant d)$. By the *set of level* $E_t$ of the function $f(x, y)$ we mean the set of all those points of $S$, at which $f(x, y) = t$,

$$E_t = E\{f(x, y) = t\}.$$

From the continuity of $f(x, y)$ it follows that all sets $E_t$ are closed. A closed set is said to be *connected* if it cannot be divided into the sum of two closed mutually disjoint sets. If a closed set is not connected, it is given in the form of the sum of a finite or infinite set of connected closed terms called *components* of this set.

We shall represent the number of components of a set of level $E_t$ by $J(t)$. Then $J(t)$ is measurable. The integral

$$\mathrm{V}(f) = \int_{-\infty}^{\infty} J(t)\, dt$$

is called the *linear variation* of the function $f(x, y)$. If $\mathrm{V}(f)$ is finite the function $f(x, y)$ is said to be *a function of bounded linear variation.*†

Moreover, $v_1(E_t)$ represents the Hausdorff length of a set of level $E_t$. This function is also measurable. Therefore the following integral has meaning

$$W(f) = \int_{-\infty}^{\infty} v_1(E_t)\, dt$$

and is called the *plane variation* of the function $f(x, y)$. The function $f(x, y)$, for which $W(f)$ is finite, is called *a function of bounded plane variation.*

The role of plane and linear variation for a function of two variables is made clear by the following theorems due to A. S. Kronrod.

† If we consider functions of one variable $f(x)$, which are continuous on $[a, b]$, and denote by $J(t)$ the number of components of the set $E_t = E\{f(x) = t\}$, then

$$\int_{-\infty}^{+\infty} J(t)\, dt = \bigvee_a^b f.$$

This equation and this transition to the study of continuous functions of bounded variation were discovered by S. Banach.

(a) *If a function $f(x, y)$, which is continuous on $S$, has bounded plane variation, then it has an asymptotic differential almost everywhere on $S$.*

(b) *If a function $f(x, y)$, which is continuous on $S$, has bounded plane and linear variation, then it possesses an exact total differential almost everywhere on $S$*

**5.** For a function of $n$ variables we may define the Lebesgue integral by exactly the same process as for a function of one variable.

Suppose that, on a bounded set $E \subset R^n$ we are given a measurable, bounded function of a point $y = f(M)$, all of whose values are included strictly between the numbers $t$ and $T$. Let us divide the segment $[t, T]$ into $l$ parts by the points $t = y_0 < y_1 < y_2 < \cdots < y_l = T$, and let us define the set

$$E_k = E\{y_{k-1} \leqslant f(M) < y_k\}.$$

Then we may construct integral sums

$$s_l = \sum_{k=1}^{l} y_{k-1} m E_k, \quad S_l = \sum_{k=1}^{l} y_k m E_k,$$

where $mE_k$ represents the $n$-dimensional Lebesgue measure of the set $E_k$. These sums are called *the lower and upper Lebesgue sums* respectively. The common limit, to which the lower and upper Lebesgue sums tend, when the number $l$ increases indefinitely, the length of all parts $[y_{k-1}, y_k]$ approaching zero, is called *the Lebesgue integral of the function $f(M)$ with respect to the set $E$.* We denote it by one of the symbols

$$\int_E f(M)\, dw \text{ or } \int_E \int \cdots \int f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n.$$

It is easy to see that this definition does not differ from that given in § 5. In this connection we shall not dwell on the properties of the Lebesgue integral and the extension of this definition to unbounded functions. Also, as for functions of one variable, the functions $f(M)$, for which the Lebesgue integral exists, are said to be *summable.*

The connection between a multiple Lebesgue integral and a one-dimensional one is established by the following theorem.

FUBINI'S THEOREM (for a function of two variables). *Suppose that we are given a summable function $f(x, y)$ in a rectangle $S(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$. Then*

(a) *for almost all $x \in [a, b]$ the function $f(x, y)$ as a function of $y$ is summable on $[c, d]$;*

(b) *if $\Delta$ denotes the set of those $x \in [a, b]$, for which $f(x, y)$ is summable on $[c, d]$, then the integral $\int_c^d f(x, y)\, dy$ as a function of $x$ is summable on $\Delta$;*

(c) *the following formula applies*

$$\iint_S f(x, y)\, dx\, dy = \int_\Delta dx \int_c^d f(x, y)\, dy.$$

If we assume that the function, which stands under the integral sign, is defined almost everywhere, then instead of $\int_\Delta$ we may write $\int_a^b$, since from (a) it follows that $m\Delta = b - a$. We then obtain the following equation

$$\iint_S f(x, y)\, dx\, dy = \int_a^b dx \int_c^d f(x,y)\, dy.$$

In view of the equivalence of the independent variables we arrive at the statement

*if $f(x, y)$ is summable on the rectangle $S$, then both repeated integrals*

$$\int_a^b dx \int_c^d f(x, y)\, dy \quad \text{and} \quad \int_c^d dy \int_a^b f(x, y)\, dx$$

*exist and*

$$\int_a^b dx \int_c^d f(x, y)\, dy = \int_c^d dy \int_a^b f(x, y)\, dx.$$

We must note that, from the existence of one or even both repeated integrals, the summability of the function $f(x, y)$ does not necessarily follow.

Similarly, Fubini's theorem may be formulated for a function of a large number of variables.

**6.** Let a bounded function of the point $f(M)$ be defined on the set $E \subset R^n$. Let us consider some measure $\mu$, i.e. a completely additive, non-negative and monotonic function of sets, which is defined on a system of sets $\sigma$, containing the set $E$. We shall call

the function $f(M)$ *measurable with respect to the measure* $\mu$, if all sets $E\{f > a\}$, for any $a$, belong to the system $\sigma$.

In such a case let us divide the part $[t, T]$, which contains all values of $f(M)$, by points $t = y_0 < y_1 < y_2 < \cdots < y_m = T$, and construct the integral sums

$$s_m = \sum_{k=1}^{m} y_{k-1}\mu(E_k) \quad \text{and} \quad S_m = \sum_{k=1}^{m} y_k\mu(E_k).$$

The common limit of the integral sums $s_m$ and $S_m$, if it exists, is called the *Lebesgue–Stieltjes integral* of the function $f(M)$ with respect to the set $E$. It is denoted by the symbol

$$\int_E f(M)\,\mu(dE).$$

From the definition it is clear that the properties of the Lebesgue–Stieltjes integral coincide with the properties of the ordinary Lebesgue integral, if we replace the Lebesgue measure (where it occurs in the formulae) by the measure $\mu$. It is obvious that after replacing the Lebesgue measure on a straight line by some other measure, we obtain the Lebesgue–Stieltjes integral for a function of one variable. The carry-over of the idea of the Lebesgue–Stieltjes integral to the case of unbounded functions also involves little difficulty.

## § 9. BASIC FUNCTION SPACES

**1.** *Metric space* is a set of objects of any nature, in which the distance between the elements is defined. This means that on a given set a function is defined, which sets in correspondence to each pair of elements (points) $x$ and $y$ some number $\varrho = \varrho(x, y)$, called the *distance* between them, which satisfies the following conditions:

(a) $\varrho(x, y) \geqslant 0$; $\varrho(x, y) = 0$ when, and only when, the elements $x$ and $y$ coincide;

(b) $\varrho(x, y) = \varrho(y, x)$;

(c) $\varrho(x, y) \leqslant \varrho(x, z) + \varrho(y, z)$.

The conditions we have enumerated are called *the axioms of the metric*. If $x$, $y$ and $z$ are points of Euclidean space and the distance between them is defined in the usual way, then the condition (c) expresses the well-known property of the triangle: the

sum of two sides of a triangle is not less than the third side. Therefore this condition is often called the *triangle inequality* or the *triangle axiom.*

As for $n$-dimensional Euclidean spaces, so for a metric space we may easily define the basic ideas considered in § 1.

A *sphere* of radius $r$ with centre at the point $x$ of a metric space $E$ is the set of elements $y \in E$, which satisfy the condition $\varrho(x, y) < r$, and the $\varepsilon$-neighbourhood of the point $x$ is the sphere of radius $\varepsilon$ with centre $x$. After this, ideas of isolated and limit points, and also of a closed and open set are easily defined. We say that a sequence $\{x_n\}$ *converges to the point* $x_0$ if $\lim_{n\to\infty} \varrho(x_n, x_0) = 0$.

The sequence $\{x_n\}$ is said to be *fundamental* if it satisfies *Cauchy's criterion*, i.e. if for any $\varepsilon > 0$ a whole number $N$ can be found such that $\varrho(x_n, x_m) < \varepsilon$ if $n \geqslant N$, $m \geqslant N$. Any convergent sequence is fundamental. If in a metric space any fundamental sequence converges then this space is said to be *complete.*

A set of elements of any nature is said to be *a linear space*, if, for its elements, operations of addition and multiplication by real numbers are defined, which satisfy the ordinary conditions.† Linear space is said to be *normed* if to each element $x \in E$ there corresponds a real number $\|x\|$, *the norm of* $x$, which satisfies the following requirements:

(a) $\|x\| \geqslant 0$; $\|x\| = 0$ when, and only when, $x$ is the null element of $E$;

(b) $\|\lambda x\| = |\lambda| \cdot \|x\|$;

(c) $\|x + y\| \leqslant \|x\| + \|y\|$.

Normed space may be made metric by introducing the distance between the elements $x$ and $y$ as the norm of their difference, $\varrho(x, y) = \|x - y\|$. Using these observations, all definitions, formulated in terms of a metric, may be formulated also in terms of a norm. For example, the sequence $\{x_n\}$ converges to $x_0$, if $\|x_n - x_0\| \to 0$. This convergence is often called *convergence in norm* or *strong convergence.*

Here we shall only consider some examples of *function spaces*, whose elements are functions of one class or other of a different

† i.e. the conditions of associativity, multiplication by a scalar, and distributivity.

form, depending on the distance between them. We shall limit ourselves to the consideration of spaces of real functions of one real variable.

**2.** The set of all continuous functions, defined on some segment, is a linear space. The distance between the elements $x$ and $y$ (functions $x(t)$ and $y(t)$) of this space is defined by the equation $\varrho(x, y) = \max |x(t) - y(t)|$. Here the function which is identically zero on the segment, serves as *the null element.* Therefore we may define its norm as $\|x\|_C = \max |x(t)|$.

The space of continuous functions with metric and norm as defined above is called *space C.* If we wish to stress the fact that we are dealing with functions defined and continuous on the segment $[a, b]$, we write $C_{[a,b]}$. In the study of Fourier trigonometric series we have to deal with a space of continuous *periodic functions,* which are represented by $CP_{[0,2\pi]}$ or $CP_{[0,1]}$, depending on the segment of definition and period.

The convergence in norm, in a space $C$, of a sequence $\{x_n\}$ to the element $x_0$ means the approach to zero of the distance $\|x_n - x_0\| = \max |x_n(t) - x_0(t)|$. Thus, *strong convergence in space C means ordinary uniform convergence.*

**3.** The set of all continuous functions, which have continuous derivatives up to order $k (k \geqslant 1)$ inclusive, is linear normed space $C^{(k)}_{[a,b]}$, if the distance between the elements $x(t)$ and $y(t)$ is defined by the equation

$$\varrho(x, y) = \max_{a \leqslant t \leqslant b} \sum_{s=0}^{k} |x^{(s)}(t) - y^{(s)}(t)|, \quad \begin{matrix} x^{(0)}(t) \equiv x(t), \\ y^{(0)}(t) \equiv y(t). \end{matrix}$$

Here

$$\|x\|_{C^{(k)}} = \max \sum_{s=0}^{k} |x^{(s)}(t)|.$$

Strong convergence in the space $C^{(k)}_{[a,b]}$ means uniform convergence of a sequence of functions and sequences which are composed of derivatives up to order $k$ inclusive.

**4.** The set of summable functions on a given segment, i.e. of measurable functions $x(t)$ for which the integral

$$\int_a^b |x(t)|\, dt$$

exists and is finite, forms *space L. The distance* between the elements $x, y \in L$ is defined by the equation

$$\varrho(x, y) = \int_a^b |x(t) - y(t)|\, dt.$$

*The null element* is here a function, which is zero almost everywhere, i.e. equivalent to identical zero. The *norm* of the element $x \in L$ may be defined as

$$\|x\|_L = \int_a^b |x(t)|\, dt.$$

In a space $L$ it is not possible to distinguish equivalent functions, because the distance between them is zero. Functions, which differ only on a set of measure zero, are considered here to be identical. Alternatively, an element of a space $L$ is not a separate summable function, but a whole class of functions, which coincide with each other almost everywhere.

Any function belonging to space $C$ is summable, i.e. it belongs also to space $L$. The inclusion $C \subset L$ is, however, correct, if we consider $C$ and $L$ simply as two sets of functions, but not true if we are talking about function spaces. In fact for one and the same function $x(t)$, belonging to both $C$ and $L$, its norm as an element of space $C$ is distinct from its norm as an element of $L$, $\|x\|_C \neq \|x\|_L$. Also the distance between two functions in spaces $C$ and $L$ will be different.

5. *To space* $L^2$, which we have already encountered in § 7, *there are related functions with a summable square*, i.e. measurable functions $x(t)$, for which the integral $\int_a^b x^2(t)\, dt$ is finite. *The distance* between the elements $x, y \in L^2$ is

$$\varrho(x, y) = \left( \int_a^b [x(t) - y(t)]^2\, dt \right)^{\frac{1}{2}};$$

a function which is zero almost everywhere serves as the *null element*, as for $L$. In accordance with this, the norm in space $L^2$

is defined as

$$\|x\|_{L^2} = \left(\int_a^b |x(t)|^2\, dt\right)^{\frac{1}{2}}.$$

Hence it is clear that *strong convergence of a sequence in space $L^2$ means convergence in the mean.*

In space $L^2$ equivalent functions are not distinct. If we consider space $L^2$ as linear and its elements as vectors, then, for $x, y \in L^2$ we may define a *scalar product* by the equation

$$(x, y) = \int_a^b x(t)y(t)\, dt.$$

Thus, elements of $L^2$ are said to be *orthogonal* (see § 7) if their scalar product is zero. Strictly also a function is *normalized* if its norm is unity, *the norm of an element playing the part of the length of a vector.*

*The Fourier series of the function $x \in L^2$ with respect to a complete orthonormal system $\{\varphi_n(x)\}$ is an analogue of the decomposition of a vector on a Cartesian basis in Euclidean space.* The Fourier coefficients are the orthogonal projections of the vector $x$ on the corresponding coordinate vectors of $\varphi_n$.

**6.** For any $p \geqslant 1$ we can define the *space $L^p$ of functions which are measurable and of degree $p$.* The distance between the elements $x, y \in L^p$ is defined by the equation

$$\varrho(x, y) = \left(\int_a^b |x(t) - y(t)|^p\, dt\right)^{\frac{1}{p}},$$

and the *norm* by the equation

$$\|x\|_{L^p} = \left(\int_a^b |x(t)|^p\, dt\right)^{\frac{1}{p}}.$$

*Strong convergence in $L^p$ means convergence in the mean of order $p$.* If $p < q$ functions belonging to $L^q$ belong also to $L^p$. Therefore if we consider $L^p$ and $L^q$ as sets of functions, the inclusion $L^q \subset L^p$ is correct. *This inclusion does not hold,* however, if we have in mind *function spaces,* because the norms of functions as elements of different spaces are different.

The spaces $L$ and $L^2$ considered in § 9, paras. 3 and 4 are particular cases of spaces $L^p$ with $p = 1$ and $p = 2$.

We must remember that in § 6, para. 4 convergence in the mean of order $p$ was defined for any exponents $p > 0$. Spaces $L^p$ with $0 < p < 1$ are, however, not generally considered because they are not metric spaces. In fact for distance

$$\varrho(x, y) = \left\{ \int_a^b |x(t) - y(t)|^p \, dt \right\}^{\frac{1}{p}},$$

with $0 < p < 1$ the inequality of the triangle is violated.

A function $x(t)$, which is defined on a segment $[a, b]$, is said to be *essentially bounded* if a number $N$ can be found such that the inequality $|x(t)| \leqslant N$ is satisfied for almost all $t$. The lower bound of such numbers $N$ is called the *essential upper bound* of the function $x(t)$ on $[a, b]$ and we write

$$\inf N = \sup_{a \leqslant t \leqslant b} \text{vrai } x(t).$$

The set of essentially bounded functions with metric

$$\varrho(x, y) = \sup_{a \leqslant t \leqslant b} \text{vrai } |x(t) - y(t)|$$

is called *the space $M$ of essentially bounded functions.*

Each of the functions, which enter into the space $M$, belong also to any of the spaces $L^p$ with any $p \geqslant 1$. As above, however, we cannot speak of the inclusion $M \subset L^p$, because for one and the same function its norms in spaces $M$ and $L^p$ will de different.

*Strong convergence in $M$ means uniform convergence* almost everywhere.

We may also consider the *space $V$ of functions of bounded variation*, if the distance is defined by the equation

$$\varrho(x, y) = \bigvee_a^b (x - y).$$

The spaces $M$ and $V$ will be normed, if for an element $x$ we define the norm by the equations

$$\|x\|_M = \sup_{a \leqslant t \leqslant b} \text{vrai } |x(t)|$$

$$\|x\|_V = \bigvee_a^b (x).$$

Let $x(t)$ be a measurable function, defined on the segment $[0, 1]$. We shall construct a set $E_x$ of such real numbers $\varepsilon$, that the set of points $t \in [0, 1]$, for which $|x(t)| > \varepsilon$, has a measure not exceeding $\varepsilon$. In any case the set $E_x$ contains the number $\varepsilon = 1$, and therefore is not empty. We now introduce the notation $\varepsilon_0(x) = \inf E_x$. The set $S$ of all functions, which are measurable on $[0, 1]$, will be a metric space, if the distance between the elements $x, y \in S$ is defined by the equation

$$\varrho(x, y) = \varepsilon_0(x - y).$$

The space $S$ is called *a space of measurable functions*. It is also called a *space of convergence in measure*, because *convergence in space $S$ means convergence in measure.*

The space $S$, as distinct from all the spaces considered, is not normed, i.e. we cannot bring the norm $\|x\|_S$ into it so that convergence in this norm would be equivalent to convergence in measure. In particular $\varepsilon_0(x)$ cannot be taken as norm, because for this function axiom (b) is violated. We must note that all the spaces we have considered, $C$, $C^{(k)}$, $L^p$, $V$, $M$, and $S$ are complete spaces.

**7.** *Spaces of sequences*, in which norms and distances are defined in a similar way to norms and distances in corresponding functional spaces, are closely connected with functional spaces. Thus, for the element $x \equiv \{x_n\}$ we have, *in a space of converging numerical sequences,*

$$\|x\|_c = \sup_{1 \leqslant n < \infty} |x_n|.$$

*In the space $l^p (p \geqslant 1)$ of sequences with a convergent sum of degree $p$, for the element*

$$x \equiv \{x_n\} \in l^p$$

$$\|x\|_{l^p} = \left( \sum_{n=1}^{\infty} |x_n|^p \right)^{\frac{1}{p}};$$

*in the space $m$ of bounded sequences*

$$\|x\|_m = \sup_{1 \leqslant n \leqslant \infty} |x_n|.$$

It must be noted that any convergent sequence is bounded, so that all sequences which belong to $c$ belong also to $m$. Moreover, sums in $c$ and $m$ coincide, and so the inclusion $c \subset m$ is valid.

A specially important part is played by the space $l^2$ with norm

$$\|x\|_{l^2} = \sqrt{\sum_{n=1}^{\infty} x_n^2}$$

and metric

$$\varrho(x, y) \sqrt{\sum_{n=1}^{\infty} (x_n - y_n)^2}.$$

From these formulae it is obvious that *the space $l^2$ is the simplest natural analogue of n-dimensional Euclidean space*, considered in § 8. The spaces $c$, $m$ and $l^p$ $(p \geqslant 1)$ are complete.

The Riesz–Fischer theorem (see § 7) establishes a unique, mutual, linear correspondence between spaces $l^2$ and $L^2$. To each function of $L^2$ there corresponds the sequence of its Fourier coefficients with respect to a given complete orthonormal system with a convergent sum of its squares, and conversely; here, to the sum of two functions and the product of a function by a number there correspond the sum of sequences and the multiplication of sequences by a number. By Parseval's equation, norms of corresponding elements from $L^2$ and $l^2$ are equal to each other, i.e. we have a *mutually unique linear transformation with conservation of distance*. Such a transformation is called *isometric*. Thus *spaces $L^2$ and $l^2$ are isometric.*

CHAPTER II

# THE INTERPOLATION AND APPROXIMATION OF FUNCTIONS

## § 1. INTRODUCTORY REMARKS

For the study of any mathematical question (either purely theoretical or practical) it is often extremely useful to substitute (with a definite degree of accuracy) for the function considered another which is in some sense a simpler function.

Suppose that on the segment $[a, b]$ of the real axis we are given a function $f(x)$ and a class $\mathfrak{G}$ of "simple" functions.

In the present chapter we shall study functions $g(x) \in \mathfrak{G}$, which in some sense (which will always be accurately given) will take certain values in common with the function $f(x)$ (interpolation of functions), or will differ "sufficiently little" from the function $f(x)$, or will be "nearest" to the function $f(x)$ in comparison with other functions of the class $\mathfrak{G}$ (approximation of functions). In the case where the defined properties of the functions of class $\mathfrak{G}$ and the character of their approximation of the function $f(x)$ have been sufficiently well studied, we may consider some properties of the function $f(x)$ itself by considering the functions of class $\mathfrak{G}$ which approximate to it. In numerical methods and calculations the given function $f(x)$ is generally replaced by definite functions of the corresponding class $\mathfrak{G}$ which are more convenient for "calculations" in the widest sense of this word, i.e. for which there are already prepared tables, programmes of calculation of their values or which are such that they can reasonably be calculated by means of machines.

Problems of this kind are met early in the study of mathematical analysis. Suppose for instance that on the interval $(a, b)$ a function $f(x)$ is given which has $n$ derivatives ($n$ is a fixed natural number) at the point $x_0 \in (a, b)$. It is required to find a polynomial $P(x)$ of degree not higher than $n$ which, together with its derivatives up to the order $n$ inclusive, coincides at the point $x_0$ with the

corresponding values of the function $f(x)$ and its derivatives:

$$P^{(k)}(x_0) = f^{(k)}(x_0), \quad k = 0, 1, \cdots, n\dagger \tag{1.1}$$

Thus in this concrete case the class $\mathfrak{G}$ consists of all polynomials of degree not higher than $n$, and that which functions of class $\mathfrak{G}$ have in common with the given function $f(x)$ is expressed by formula (1.1). As is well known, such a polynomial $P(x)$ always exists, is unique and forms the principal part of Taylor's formula of order $n$ at the point $x_0$, i.e.

$$P(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n. \tag{1.2}$$

It must be noted also that the polynomial (1.2) gives a solution of one more important problem, viz. from among all polynomials of degree not higher than $n$ it best approximates the given function $f(x)$ as $x$ approaches $x_0$. This is expressed in more detail and more accurately as follows: on the one hand, for the polynomial $P(x)$ given by the formula (1.2) we have the equation

$$f(x) - P(x) = o((x - x_0)^n), \quad \text{as } x \to x_0, \tag{1.3}$$

but, on the other hand, there exists no polynomial of degree higher than $n$ different from the polynomial (1.2) which possesses the property (1.3).

## § 2. THE INTERPOLATION OF FUNCTIONS BY POLYNOMIALS

**2.1. The simplest interpolation problem. The Lagrange and Newtonian interpolation polynomials.** Suppose that on the segment $[a, b]$ we are given a real function $f(x)$ and $m + 1$ fixed values of the argument $x_i$, $i = 1, 2 \ldots, m, m + 1$,

$$x_1 < x_2 < \cdots < x_{m+1}. \tag{2.1}$$

One of the simplest interpolation problems is that of finding a polynomial $P(x)$, not higher than some degree $n$, which takes on the same value as the given function for given values of the argument (the points of interpolation), i.e. for which

$$f(x_k) = P(x_k), \quad k = 1, 2, \ldots, m + 1. \tag{2.2}$$

† The symbol $\varphi^{(0)}(x)$ denotes as usual the actual function $\varphi(x)$ being considered.

Such a polynomial $P(x)$ is said to be an *interpolation polynomial*, interpolating the function $f(x)$ at the given points of interpolation.

From now on the aggregate of all polynomials of degree not higher than $n$, i.e. the aggregate of all functions of the form

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n, \tag{2.3}$$

where the $a_i$ are real numbers, $i = 0, 1, 2, \ldots, n$, will be represented by $\mathfrak{P}_n$, $n = 0, 1, 2, \ldots$

To consider the question of the existence of the polynomial $P(x) \in \mathfrak{P}_n$, which satisfies conditions (2.2), we shall take a polynomial $P(x) \in \mathfrak{P}_n$ of the form (2.3) with indeterminate coefficients $a_i$, $i = 0, 1, \ldots, n$, and substitute it in the system (2.2). We shall obtain a system of $m + 1$ linear equations with $n + 1$ unknowns $a_0, a_1, \ldots, a_n$:

$$\left.\begin{aligned} a_0 + a_1 x_1 + \cdots + a_n x_1^n &= f(x_1), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_0 + a_1 x_{m+1} + \cdots + a_n x_{m+1}^n &= f(x_{m+1}). \end{aligned}\right\} \tag{2.4}$$

Note that the determinant, composed of the coefficients of this system in the first $k$ rows and the first $k$ columns, is Vandermonde's determinant,† $W(x_1, \ldots, x_k)$ which in this case is not zero because all the points of interpolation are different. Therefore the rank of the matrix of the system (2.4) is the lesser of the two numbers $m + 1$, and $n + 1$ ($m + 1$ is the number of rows and $n + 1$ the number of columns of this matrix). If $m > n$ the system (2.4), generally speaking, has no solution. If $m = n$ the solution of the system (2.4) always exists and is also unique. If, however, $m < n$ there exists a non-determinate set of solutions of system (2.4). Thus the minimum natural number $n$, for which the interpolation problem (2.2) can be solved with the condition that $P(x) \in \mathfrak{P}_n$, must be not less than the number of points of interpolation decreased by unity. i.e. we must have $n \geqslant m$.

The given method obviously allows us not only to prove the existence (and uniqueness in the case where $m = n$) of the interpolation polynomial, but also in fact to find it; for this it is sufficient to solve system (2.4).

We shall denote by $P(x)$ the interpolation polynomial which satisfies the system of equations (2.2). It may be shown (e.g. by solving the system (2.4)) that if $n = m$ this polynomial, in this

† See Mathematical Reference Library, *Higher algebra*, ch. 1.

case the so-called *Lagrange interpolation polynomial*, may be written in the form

$$P(x) = \sum_{k=0}^{n} f(x_k) \frac{(x - x_1)(x - x_2) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_{n+1})}{(x_k - x_1)(x_k - x_2) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_{n+1})}. \tag{2.5}$$

Also, that the written expression is a polynomial of degree not higher than $n$ and satisfies the condition (2.2) is easily verified directly.

Suppose now that derivatives of the function $f(x)$ exist on the segment $[a, b]$ up to the order $n + 1$ inclusive, then there exists a point $\xi \in (a, b)$ such that

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi)}{(n + 1)!} (x - x_1)(x - x_2) \cdots (x - x_{n+1}),$$

where $P(x)$ is the interpolation polynomial (2.5).

Assuming for short that

$$\omega(x) = (x - x_1)(x - x_2) \cdots (x - x_{n+1}),$$

$$R(x) = \frac{f^{(n+1)}(\xi)}{(n + 1)!} \omega(x)$$

and

$$M_{n+1}(a, b) = \sup_{a \leqslant x \leqslant b} |f^{n+1}(x)|,$$

we have a value for the error with the interpolation of the function $f$ from the formula (2.5)

$$|R(x)| \leqslant \frac{M_{n+1}(a, b)}{(n + 1)!} \max_{a \leqslant x \leqslant b} |\omega(x)|.$$

From this formula it is obvious that if we wish to diminish the error with the interpolation considered, then it is sufficient to choose the points of interpolation $x_1, \ldots, x_{n+1}$, on the segment $[a, b]$ so that the value $\max_{a \leqslant x \leqslant b} |\omega(x)|$ is diminished. Note that in the case where $a = -1$, $b = 1$ the polynomials $\omega(x)$, for which $\max_{a \leqslant x \leqslant b} |\omega(x)|$ reach the least value, are called *Chebyshev polynomials* (see § 3.4); they may be written in the following form:

$$T_{n+1}(x) = \frac{1}{2^n} \cos \{(n + 1) \arccos x\}, \quad n = 0, 1, 2, \cdots \dagger$$

† It must be borne in mind that polynomials $T_n(x) = \cos(n \arccos x)$, $n = 1, 2, \ldots$ are often called Chebyshev polynomials.

It must be noted that the roots of the Chebyshev polynomial $T_n(k)$ are simple, lie in the interval $(-1, 1)$ and are given by the formula

$$x_{k,n} = \cos \frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \ldots, n; \quad n = 1, 2, \ldots$$

Between two neighbouring roots $x_{k,n}$ and $x_{k+1,n}$ of the polynomial $T_n(x)$ may be found one and only one root of the polynomial $T_{n-1}(x)$, $n = 1, 2, \ldots$

This makes clear the fact that, for interpolation on the segment $[-1, 1]$, it turns out to be advantageous to take points $x_{kn}$ as points of interpolation, i.e. zeros of the Chebyshev polynomial $T_{n+1}(x)$ because here $\omega(x) = T_n(x)$. In the case of an arbitrary segment $[a, b]$ from the same point of view it is advantageous to take as points of interpolation points which are obtained from the zeros of the corresponding Chebyshev polynomial, by similar mapping of the segment $[-1, +1]$ on the segment $[a, b]$ (on interpolation with Chebyshev points see also § 2.4). For more details of Chebyshev polynomials see the Mathematical Reference Library, *Mathematical analysis (functions, limits, series, continued fractions)*, ch. X, § 15.

We shall now give another form of description of the Lagrange interpolation polynomial $P(x)$ for the simple case of equidistant points, i.e. when $x_i - x_{i-1} = \dfrac{b-a}{n}$, $i = 2, 3, \ldots, n+1$, $x_1 = a$, $x_{n+1} = b$. For this we must previously bear in mind the idea of the differences $\Delta_n^{(k)}(x, f)$ of the function $f(x)$ with step $h$. Let $h$ be some fixed number; then by definition

$$\Delta_h^{(0)}(x, f) = f(x),$$
$$\Delta_h(x, f) = \Delta_h^{(1)}(x, f) = f(x+h) - f(x),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\Delta_h^{(k+1)}(x, f) = \Delta_h\left(x, \Delta_h^{(k)}(x, f)\right) = \sum_{j=0}^{k+1} (-1)^{k-j+1} C_{k+1}^j f(x + jh).$$

The last formula is proved by induction. Note that also, conversely, the values of the function $f(x)$ at the points $x + kh$ $(k = 0, 1, \ldots, n)$ are expressed as linear combinations of successive finite differences at the initial point $x$:

$$f(x + kh) = \sum_{j=0}^{k} C_k^j \Delta_h^{(j)}(x, f), \quad k = 0, 1, \ldots, n.$$

Suppose now that on the segment $[a, b]$ we choose equidistant points of interpolation:

$$x_k = a + (k-1)h, \quad k = 1, \ldots, n+1, \quad h = \frac{b-a}{n}.$$

Then the interpolation polynomial $P(x)$, which satisfies the conditions (2.2) (with $m = n$), may be written by means of differences in the form

$$P(x) = \sum_{k=1}^{n} \frac{\Delta_h^{(k)}(a, f)(x-a)(x-a-h)\cdots[x-a-(k-1)h]}{h^k k!}. \tag{2.6}$$

The interpolation polynomial, written in the form (2.6), is called *Newton's interpolation polynomial.*

In the case of equidistant points of interpolation Newton's interpolation polynomial turns out to be in general more convenient for calculation than that of Lagrange. This is due in particular to the fact that if one point of interpolation is added, while retaining the former ones (then the segment on which the interpolation is performed is increased by $h$), and the degree of the interpolation is increased by one, then in formula (2.6) just one term is added and also the Lagrange interpolation polynomial in this case must be completely calculated anew. Next also in the case of an arbitrary distribution of points the interpolation polynomial $P(x)$ may be given such a form that the given property will be retained (see, for example [7], pp. 60–62).

A natural extension of the problem considered is the following: suppose that on the segment $[a, b]$ a system of $n$ linearly independent continuous functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ is given. As functions, with whose help we shall carry out the interpolation, we shall take functions of the form

$$\varphi(x) = \lambda_1\varphi_1(x) + \cdots + \lambda_n\varphi_n(x).$$

The interpolation problem for the given function $f(x)$, defined on the segment $[a, b]$, with points of interpolation $x_1, \ldots, x_n$, consists of finding numbers $\lambda_1, \ldots, \lambda_n$, such that

$$f(x_i) = \varphi(x_i), \quad i = 1, 2, \ldots, n.$$

In order that this interpolation problem may have a solution, and that a unique one, for any function $f(x)$ it is necessary and suf-

ficient that the determinant

$$D(\varphi_1, \cdots, \varphi_n) = \begin{vmatrix} \varphi_1(x_1)\, \varphi_2(x_1) \cdots \varphi_n(x_1), \\ \varphi_1(x_2)\, \varphi_2(x_2) \cdots \varphi_n(x_2), \\ \cdots\cdots\cdots\cdots\cdots \\ \varphi_1(x_n)\, \varphi_2(x_n) \cdots \varphi_n(x_n) \end{vmatrix}$$

should be different from zero in the case where all the $x_1, \ldots, x_n$, are mutually different. This condition is equivalent to the condition that any "generalized polynomial"

$$\varphi(x) = \lambda_1\varphi_1(x) + \cdots + \lambda_n\varphi_n(x),$$

not identically equal to zero, should vanish on the segment $[a, b]$ at not more than $n - 1$ distinct points. Such systems of functions are called Chebyshevian (see § 7.3).

Since the determinant $D(1, x, x^2, \ldots, x^{n-1})$ is Vandermonde's determinant, it satisfies the given requirement and so the system of powers $1, x, \ldots, x^{n-1}$ is Chebyshevian.

In the following paragraph we shall consider the problem of interpolation of functions in the case where the functions $\varphi_i(x)$ are trigonometric or more accurately, sines and cosines of multiple arcs.

**2.2. Interpolation by trigonometric polynomials.** A function of the form

$$T(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx), \quad a_n^2 + b_n^2 < 0, \quad (2.7)$$

is called *a trigonometric polynomial of degree n.*

With the help of complex numbers it may be written in the form †

$$T(x) = \sum_{k=-n}^{n} c_k e^{ikx}, \quad \overline{c_k} = c_{-k}, \quad k = 0, \pm 1, \pm 2, \ldots, \pm n.$$

We shall represent the aggregate of all trigonometric polynomials of degree not higher than $n$ by $\mathfrak{T}_n$.

Now let a periodic function $f(x)$ of period $2\pi$ ‡ be given and let there be fixed points of interpolation $x_0, x_1, \ldots, x_{2m}$, where $x_{j_1} - x_{j_2} \neq 2k\pi$ if $j_1 \neq j_2$, where $k$ is an integer.

† Remembering that $e^{it} = \cos t + i \sin t$.

‡ In the case of periodic functions with period $2l$, in the present section and further on, for formulating the corresponding results we need only make the linear change of variable $x = \frac{\pi}{l} t$.

A trigonometric polynomial $T(x)$ such that

$$T(x_j) = f(x_j), \quad j = 0, 1, \ldots, 2m, \tag{2.8}$$

is called an *interpolation polynomial*, interpolating the function $f(x)$ at the points $x_j$ $(j = 0, 1, \ldots, 2m)$. If $n \geqslant m$ there always exists in the class $\mathfrak{T}_n$ an interpolation polynomial which satisfies the condition (2.8), while in the case where $m = n$ it is unique and may be written in the form

$$T(x) = \sum_{k=0}^{2n} \frac{\sin\frac{x-x_0}{2} \cdots \sin\frac{x-x_{k-1}}{2} \sin\frac{x-x_{k+1}}{2} \cdots \sin\frac{x-x_{2n}}{2}}{\sin\frac{x_k-x_0}{2} \cdots \sin\frac{x_k-x_{k-1}}{2} \sin\frac{x_k-x_{k+1}}{2} \cdots \sin\frac{x_k-x_{2n}}{2}} f(x_k),$$

$$T(x) \in \mathfrak{T}_n. \tag{2.9}$$

Sometimes interpolation with the help of trigonometric polynomials which do not contain sines or cosines of multiple arcs turns out to be useful, i.e. either with the help of polynomials of the form

$$T(x) = a_0 + a_1 \cos x + \cdots + a_n \cos nx, \tag{2.10}$$

which are obviously even functions, or with polynomials of the form

$$T(x) = b_1 \sin x + \cdots + b_n \sin nx, \tag{2.11}$$

which are obviously odd functions. For the solvability of the interpolation problem by means of polynomials of the form (2.10) we must be given $m + 1$ points of interpolation $x_j$, $j = 0, 1, \ldots, m$, $m \leqslant n$, where $x_{j_1} \pm x_{j_2} \neq 2p\pi$, $j_1 \neq j_2$ and $p$ is an integer. If $m = n$ the trigonometric polynomial of the form (2.10) satisfying the conditions

$$T(x_j) = f(x_j), \quad j = 0, 1, \ldots, n,$$

is unique and may be written in the form

$$T(x) = \sum_{k=0}^{n} \frac{(\cos x - \cos x_0) \cdots (\cos x - \cos x_{k-1})(\cos x - \cos x_{k+1}) \cdots (\cos x - \cos x_n)}{(\cos x_k - \cos x_0) \cdots (\cos x_k - \cos x_{k-1})(\cos x_k - \cos x_{k+1}) \cdots (\cos x_k - \cos x_n)} f(x_k),$$

Similarly for the solvability of the interpolation problem by means of polynomials of the form (2.11) we must be given $m$ points of interpolation $x_j$, $j = 1, 2, \ldots, m$, $x_{j_1} \pm x_{j_2} \neq 2p\pi$, $j_1 \neq j_2$, $p$ is an integer, where again $m \leqslant n$. If $m = n$, the trigonometric

polynomial of the form (2.11) which satisfies

$$T(x_j) = f(x_j) \quad j = 1, 2, \ldots, n,$$

is unique and may be written in the form

$$T(x) = \sum_{k=1}^{n} \frac{\sin x\,(\cos x - \cos x_1)\cdots \cdots (\cos x - \cos x_{k-1})(\cos x - \cos x_{k+1})\cdots(\cos x - \cos x_n)}{\sin x_k\,(\cos x_k - \cos x_1)\cdots \cdots (\cos x_k - \cos x_{k-1})(\cos x_k - \cos x_{k+1})\cdots(\cos x_k - \cos x_n)} f(x_k).$$

From what has been said it is obvious that with interpolation by trigonometric polynomials, as with interpolation with algebraic polynomials, the problem of the interpolation of functions has a unique solution in the case where the number of points of interpolation coincides with the number of coefficients of the required interpolation polynomial.

We must also note the case of equidistant points of interpolation

$$x_k = \frac{2j\pi}{2n+1}, \; j = 0, 1, 2, \ldots, 2n.$$

In this case the trigonometric polynomial, which satisfies the conditions (2.8), if $m = n$, may be written in the form

$$T(x) = \frac{1}{2n+1} \sum_{j=0}^{2n} f(x_k) \frac{\sin \dfrac{2n+1}{2}(x - x_j)}{\sin \dfrac{x - x_j}{2}}.$$

If it is written in the "canonical form" (2.7)

$$T(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

then for coefficients $a_k$ and $b_k$ we have the formulae

$$a_0 = \frac{1}{2n+1} \sum_{l=0}^{2n} f(x_l),$$

$$a_k = \frac{2}{2n+1} \sum_{l=0}^{2n} f(x_l) \cos kx_l,$$

$$b_k = \frac{2}{2n+1} \sum_{l=0}^{2n} f(x_l) \sin kx_l.$$

**2.3. General interpolation problems. Hermite's formula.** For the interpolation of a function $f(x)$ by polynomials with fixed points of interpolation we may require that, at one or other of the points, not only should the values of the function coincide with the values of the interpolation polynomial, but also that the values of some derivatives of $f(x)$ should coincide with the respective values of the derivatives of the given polynomial. We shall consider one of the problems of this type.

Suppose we are given $s$ points of interpolation $x_1, \ldots, x_s$, a natural number $n$, and $s$ natural numbers $\alpha_1, \ldots, \alpha_s$ such that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_s = n + 1.$$

It is required to find a polynomial $P(x)$ of the lowest possible degree such that

$$P^{(m)}(x_k) = f^{(m)}(x_k), \qquad (2.12)$$
$$m = 0, 1, \ldots, \alpha_k - 1, \quad k = 1, 2, \ldots, s.$$

If some $\alpha_i > 1$ the corresponding point $x_i$ is called a *multiple point* of interpolation of multiplicity $\alpha_i$ (if $\alpha_i = 1$ the point $x_i$ is said to be *simple*). The formulated problem (generally called *Hermite's problem*) is obviously a direct generalization of the problem which we have considered in § 2.1, which is obtained from that with $\alpha_1 = \ldots = \alpha_s = 1$, $s = n + 1$.

Hermite's problem always has a solution and this is unique. To write down the formula for this solution we shall introduce the notation

$$\Omega(x) = \prod_{\nu=1}^{s} (x - x_\nu)^{\alpha_\nu}. \qquad (2.13)$$

Let us also expand the rational function $\dfrac{1}{k!} \dfrac{(x - x_i)^{\alpha_i}}{\Omega(x)}$ which is regular in the neighbourhood of $x_i$, in a Taylor series and denote by $l_{ik}(x)$ the sum of those terms of the expansion, the degrees of whose factors $x - x_i$ do not exceed $\alpha_i - k - 1$. Obviously $l_{ik}(x)$ is a polynomial of degree not higher than $\alpha_i - k - 1$.

The polynomial of lowest degree which satisfies the conditions (2.12) has a degree not higher than $n$ and is generally called a *Hermite interpolation polynomial*. It may be written in the form

$$P(x) = \sum_{i=1}^{s} \left[ \frac{\Omega(x)}{(x - x_i)^{\alpha_i}} \sum_{k=0}^{\alpha_i - 1} f^{(k)}(x_i) (x - x_i)^k l_{ik}(x) \right]. \qquad (2.14)$$

Other ways of writing this polynomial may be found, for example, in [4].

Note the particular case of this formula with $\alpha_1 = \alpha_2 = \ldots = \alpha_s = 2$. Assuming that

$$\omega(x) \prod_{\nu=1}^{s} (x - x_\nu), \tag{2.15}$$

$$A_k(x) = \left[1 - \frac{\omega''(x_k)}{\omega'(x_k)} (x - x_k)\right] \frac{\omega^2(x)}{\omega'^2(x)(x - x_k)^2},$$

$$B_k(x) = (x - x_k) \frac{\omega^2(x)}{\omega'^2(x)(x - x_k)^2},$$

we obtain

$$P(x) = \sum_{k=1}^{s} f(x_k) A_k(x) + \sum_{k=1}^{s} f'(x_k) B_k(x).$$

Returning to Hermite's problem in the general case we note the writing of the interpolation formula with a remainder term for the function $f(x)$, which is defined together with its derivatives up to the order $n + 1$ inclusive on the segment $[a, b]$

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} \Omega(x), \quad a < \xi < b.$$

Here $P(x)$ is defined by the formula (2.14) and $\Omega(x)$ by formula (2.13).

**2.4. The convergence of interpolation polynomials.** In the present section we shall consider the question of the convergence of sequences of interpolation polynomials to the interpolated function as the number of points increases. The simplest answer is obtained for integral functions.†

THEOREM 1. *Let the function $f(x)$ be integral. Also let some finite segment $[a, b]$ be fixed, on which for each $k = 1, 2, \ldots$, points of interpolation $x_{1k}, \ldots, x_{n_k k}$ of multiplicity $\alpha_{1k}, \ldots, \alpha_{n_k k}$ respectively are given, and let $P_k(x)$ be an interpolation polynomial such that*

$$P_k^{(i)}(x_{jk}) = f^{(i)}(x_{jk}),$$

$$i = 0, 1, \ldots, \alpha_{jk} - 1, \quad j = 1, 2, \ldots, n_k, \quad k = 1, 2, \ldots$$

† It must be borne in mind that a function $f(x)$ is said to be integral if it can be expanded in a power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$, which converges with any values of $x$.

*Then if as* $k \to \infty$ *the number of points of interpolation increases without limit, i.e.* $n_k \to \infty$, *the sequence* $P_k(x)$ *converges uniformly on the segment* $[a, b]$ *to the function* $f(x)$.

Note that for a function, about which it is only known that it is continuous, the convergence of the sequence of its interpolation polynomials to it, with the increase of the number of points, cannot be guaranteed in general. As an example of a similar assertion we have the following theorem.

THEOREM 2 (S. N. Bernstein). *Let* $P_n(x)$ *be an interpolation polynomial which interpolates the function* $|x|$ *on the segment* $[-1, 1]$ *at* $n$ *equidistant points (see §. 2.1). Then the sequence of polynomials* $P_n(x)$ *does not converge to the function* $|x|$ *at any point of the segment* $[-1, 1]$ *besides the points* $-1, 0, 1$, *at which it converges to the corresponding values of* $|x|$.

Other negative results related to the question of the convergence of interpolation processes may be found, for example, in [7] and [11]. We now consider, on the contrary, some of the positive results in this direction.

Suppose that on the segment $[a, b]$ we are given $n$ points

$$x_1, \ldots, x_n. \tag{2.16}$$

Let us assume that

$$l_k(x) = \frac{\omega(x)}{\omega'(x_k)\,(x - x_k)}, \quad k = 1, 2, \ldots, n,\dagger$$

$$\lambda_n(x) = \sum_{k=1}^{n} |l_k(x)|,$$

$$\lambda_n = \max_{a \leqslant x \leqslant b} \lambda_n(x),$$

$l_k(x)$ are polynomials of degree $n - 1$ such that

$$l_k(x_i) = \begin{cases} 0, & \text{if } i \neq k, \\ 1, & \text{if } i = k. \end{cases} \tag{2.17}$$

The polynomials $l_k(x)$ are generally called *fundamental* with respect to the system of $n$ points. They are uniquely defined by the system (2.16) and the property (2.17).

It is obvious that the Lagrange interpolation polynomial $P(x)$ with points of interpolation $x_1, \ldots, x_n$ in the notation of (2.17) takes the form

$$P(x) = \sum_{k=1}^{n} f(x_k) l_k(x). \tag{2.18}$$

† The symbol $\omega(x)$ is defined in (2.15).

It must be noted that the character of the behaviour of the function $\lambda_n(x)$ plays an essential part in questions of convergence of interpolation processes.

To formulate further results we shall give one more definition: let the function $f(x)$ be defined on the segment $[a, b]$. The quantity

$$E_{\mathfrak{P}_n}(f) = \inf_{P(x) \in \mathfrak{P}_n} \sup_{x \in [a, b]} |f(x) - P(x)|$$

is called the *best approximation* of the function $f(x)$ by polynomials of degree not higher than $n$.†

The class of functions which are defined and continuous on the segment $[a, b]$, together with their derivatives up to order $k$ inclusive ($k = 0, 1, 2, \ldots$), we shall represent by $C_k[a, b]$. If $k = 0$ we shall simply write $C[a, b]$ instead of $C^0[a, b]$.

THEOREM 3. *Let* $f(x) \in C[a, b]$, $x_0 \in [a, b]$ *and let*

$$\lim_{n \to \infty} \lambda_n(x_0) E_{\mathfrak{P}_{n-1}}(f) = 0.$$

*Then if the Lagrange interpolation polynomials* $P_n(x) \in \mathfrak{P}_{n-1}$, $n = 1, 2, \ldots$, *interpolate the function* $f(x)$ *at the points* (2.16)

$$\lim_{n \to \infty} P_n(x_0) = f(x_0).$$

THEOREM 4. *Let* $f(x) \in C[a, b]$, $D$ *be some subset of the segment* $[a, b]$, $\mu_n = \sup_{x \in D} \lambda_n(x)$ *and let*

$$\lim_{n \to \infty} \mu_n E_{\mathfrak{P}_{n-1}}(f) = 0,$$

*then the sequence of Lagrange interpolation polynomials* $P_n(x) \in \mathfrak{P}_{n-1}$, *interpolating the function* $f(x)$ *at the points* (2.16), *converges uniformly on the set* $D$ *to the function* $f(x)$.

COROLLARY. Let the function $f(x) \in C[a, b]$ and

$$\lim_{n \to \infty} \lambda_n E_{\mathfrak{P}_{n-1}}(f) = 0;$$

then the sequence of Lagrange interpolation polynomials $P_n(x) \in \mathfrak{P}_{n-1}$, which interpolate the function $f(x)$ at the points (2.16), converges uniformly on the segment $[a, b]$ to the function $f(x)$.

Thus, roughly speaking, the slower the growth of the quantity $\lambda_n(x)$ (correspondingly $\lambda_n$) the wider the class of functions for which there is convergence of the interpolation process.

† For a more detailed account of best approximations, see § 3.

The growth of the quantity $\lambda_n$ is evaluated as follows; with any choice of points of interpolation (2.16) we have the inequality (S. N. Bernstein and G. Faber)

$$\lambda_n > \frac{\ln n}{8\sqrt{\pi}}.$$

On the other hand if, in the case of the segment $[-1, 1]$ for the points of interpolation we take the Chebyshev points (see § 2.1)

$$x_k = \cos \frac{2k+1}{2n}\pi, \quad k = 0, 1, \ldots, n-1, \tag{2.19}$$

then (S. N. Bernstein)

$$\lambda_n \leqslant 8 + \frac{4}{\pi} \ln n, \quad n = 1, 2, \ldots$$

These two inequalities show that if as points of interpolation we take the Chebyshev points we shall obtain the lowest order of growth of the quantity $\lambda_n$.

Using the results of the theory of best approximations of functions (see § 3), from those formulated above we can obtain various criteria for the convergence of the interpolation processes. We must previously have in mind one more important idea.

For any function $f(x)$ which is defined on some set $E$, the function

$$\omega(\delta; f) = \sup_{\varrho(x', x'') \leqslant \delta} [f(x') - f(x'')], \; x' \in E, \; x'' \in E \dagger$$

defined for any $\delta \geqslant 0$, is called the *modulus of continuity* of the function $f(x)$.

THEOREM 5. *Let the function $f(x)$ be defined on the segment $[-1, 1]$ and satisfy the condition*

$$\lim_{\delta \to 0} \omega(\delta; f) \ln \delta = 0$$

(this is generally called a *Dini-Lipschitz condition*); *then the sequence of Lagrange interpolation polynomials $P_n(x) \in \mathfrak{P}_{n-1}$, which interpolate the function $f(x)$ at the Chebyshev points* (2.19), *converges uniformly on the segment $[-1, 1]$ to the function $f(x)$.*

Theorems similar to those already formulated hold also in the case of interpolation by trigonometric polynomials; corresponding results may be found, for example, in [11].

† The symbol $\varrho(x', x'')$ denotes the distance between the points $x'$ and $x''$ and so the given definition of the modulus of continuity makes sense in the case where $E$ is a metric space.

We note also two more theorems similar in character.

THEOREM 6 (V. I. Krylov). *Let the function $f(x)$ be absolutely continuous on the segment $[-1, 1]$. Then the sequence of Lagrange interpolation polynomials $P_n(x) \in \mathfrak{P}_{n-1}$, which interpolate the function $f(x)$ at the Chebyshev points* (2.19), *converges uniformly on the segment $[-1, 1]$ to the function $f(x)$.*

THEOREM 7 (V. I. Krylov). *Let the function $f(x)$ be defined on the segment $[-1, 1]$ and have bounded variation. Then the sequence of Lagrange interpolation polynomials $P_n(x) \in \mathfrak{P}_{n-1}$, which interpolate the function $f(x)$ at the Chebyshev points* (2.19), *converges at each point of continuity of the function $f(x)$ to the value of the function $f(x)$ at that point.*

Now suppose that on the segment $[a, b]$ we are given a non-negative, summable function $q(x)$ which is not zero almost everywhere. We shall in future call this function *a weight function* or, more briefly, simply *a weight*. Suppose further that $\{p_n(x)\}$ is some system of polynomials which is orthogonal on the segment $[a, b]$ with weight $q(x)$:

$$\int_a^b q(x)\, p_n(x)\, p_m(x)\, dx = 0 \quad \text{if} \quad n \neq m.$$

Finally, let

$$x_{1n}, x_{2n}, \ldots, x_{nn} \tag{2.20}$$

be the zeros of the polynomial $p_n(x)$, $n = 1, 2, \ldots$

If we take the system (2.20) as points of interpolation, then in a sense, as was done for interpolation at the Chebyshev points,† in this case also we are able to obtain a series of criteria for the convergence of the interpolation processes; some of the related results are put forward, for example, in [11].

Here we note only one of the most general results, related it is true to the convergence in the mean of interpolation polynomials.

THEOREM 8 (R. Erdös and R. Turan). *Let $f(x) \in C[a, b]$, and let the Lagrange interpolation polynomial $P_n(x) \in \mathfrak{P}_{n-1}$ interpolate the function $f(x)$ at the points* (2.20); *then*

$$\lim_{n\to\infty} \int_a^b q(x)\, [P_n(x) - f(x)]^2\, dx = 0,$$

† We must remember that Chebyshev polynomials form an orthogonal system on the segment $[-1, 1]$ with weight $\dfrac{1}{\sqrt{1 - x^2}}$.

i.e. *the sequence of polynomials* $\{P_n(x)\}$ *converges in the mean on the segment* $[a, b]$ *with weight* $q(x)$ *to the function* $f(x)$.

Another way of studying the convergence of interpolation processes, which is suitable for investigation also in the case of multiple roots, is related to the idea of the so-called normal matrix of the points of interpolation.

Suppose we are given a system of points of interpolation, which we can write in the form of a triangular matrix

$$\begin{pmatrix} x_{11} & & & \\ x_{12} & x_{22} & & \\ \cdots & \cdots & \cdots & \\ x_{1n} & x_{2n} & \cdots & x_{nn} \\ \cdots & \cdots & \cdots & \end{pmatrix}, \tag{2.21}$$

where in the $n$th row are written the points of interpolation $x_{kn} \in [a, b]$, $k = 1, 2, \ldots, n$, $n = 1, 2, \ldots$, for the interpolation of the function by means of the polynomial $P(x) \in \mathfrak{P}_{n-1}$. Let us consider the functions

$$v_{kn}(x) = 1 - \frac{\omega_n''(x_{kn})}{\omega_n'(x_{kn})}(x - x_{kn}), \quad k = 1, 2, \ldots \dagger$$

where

$$\omega_n(x) = \prod_{k=1}^{n}(x - x_{kn}).$$

If for all $x \in [a, b]$

$$v_{kn}(x) \geqslant 0, \quad k = 1, 2, \ldots, n, \quad n = 1, 2, \ldots,$$

the matrix (2.21) is said to be *normal*; but if there exists $\sigma > 0$ such that for all $x \in [a, b]$

$$v_{kn}(x) \geqslant \sigma > 0, \quad k = 1, 2, \ldots, n, \quad n = 1, 2, \ldots,$$

the matrix (2.21) is said to be *strictly normal.*

As an example we note that if $x_{kn}$ $(k = 1, 2, \ldots, n, \quad n = 1, 2, \ldots)$ are the roots of the Jacobi an polynomials $J_n^{(\alpha,\beta)}(x)$,‡ then with $\alpha \leqslant 0$, $\beta \leqslant 0$ the matrix (2.21) is normal and if $\alpha < 0$, $\beta < 0$ it is strictly normal.

THEOREM 9 (L. Féjer). *Let the function* $f(x)$ *be defined on the segment* $[a, b]$ *and satisfy Hölder's condition with exponent* $\alpha > \frac{1}{2}$

$$|f(x') - f(x'')| < M\,|x' - x''|^{\alpha},$$

*where* $M$ *is constant,* $x' \in [a, b]$, $x'' \in [a, b]$.

† Functions $v_{kn}(x)$ appear in the solution of Hermite's problem with double points, see (2.15).

‡ Mathematical Reference Library, *Mathematical analysis, (Functions, limits, series, continued fractions)*, ch. IV, § 4.

*Also let the matrix of the points of interpolation* (2.21) *be normal and* $P_n(x)$ *be a sequence of Lagrange interpolation polynomials, which interpolate the function* $f(x)$ *at the corresponding points of the matrix* (2.21). *Then for any* $h > 0$ *the sequence* $\{P_n(x)\}$ *converges uniformly as* $n \to \infty$ *to the function* $f(x)$ *on the segment* $[a + h, b - h]$. *If the matrix* (2.21) *is strictly normal the sequence* $\{P_n(x)\}$ *converges uniformly to the function* $f(x)$ *on the whole segment* $[a, b]$.

THEOREM 10 (L. Féjer). *Let* $f(x) \in C'[a, b]$, *the matrix of the points of interpolation* (2.21) *be normal, and* $Q_n(x) \in \mathfrak{P}_{2n-1}$ *be a Hermite interpolation polynomial satisfying the condition*

$$Q_n(x_{kn}) = f(x_{kn}), \quad Q_n'(x_{kn}) = f'(x_{kn}),$$
$$k = 1, 2, \ldots, n, \quad n = 1, 2, \ldots$$

*Then for any* $h > 0$ *the sequence* $\{Q_n(x)\}$ *converges uniformly to the function* $f(x)$ *on the segment* $(a + h, b - h]$. *If the matrix* (2.21) *is strictly normal then the sequence* $\{Q_n(x)\}$ *converges uniformly to the function* $f(x)$ *on the whole segment* $[a, b]$.

We have considered some of the criteria for the convergence of Lagrange and Hermite interpolation polynomials connected with some limitation or other on the class of interpolation functions. In conclusion it must be noted that methods of interpolation exist which ensure the convergence of the respective interpolation polynomials for any previously given continuous function. These methods are connected with the introduction of interpolation polynomials of sufficiently large degree in comparison with the number of points of interpolation. As an example we have the following theorem.

THEOREM 11 (L. Féjer). *Let*

$$x_{kn} = \cos \frac{2k + 1}{2n} \pi, \quad k = 0, 1, \ldots, n - 1, \quad n = 1, 2, \ldots$$

(*i.e.* $x_{kn}$ *are Chebyshev points*) *and the function* $f(x) \in C[-1, 1]$. *Then if* $Q_n(x) \in \mathfrak{P}_{2n-1}$, *for any* $n = 1, 2, \ldots$, *is an interpolation polynomial such that*

$$Q_n(x_{kn}) = f(x_{kn}), \quad Q_n'(x_{kn}) = 0, \quad k = 0, 1, 2, \cdots, n - 1,\dagger$$

*on the segment* $[-1, 1]$ *the sequence* $\{Q_n(x)\}$ *converges uniformly to the function* $f(x)$.

† Such a polynomial $Q_n(x)$ always exists.

It was also shown by L. Féjer that the condition for the derivatives of the interpolation polynomials $Q_n(x)$ to vanish at the Chebyshev points may be essentially relaxed, i.e. it is sufficient that the derivatives $Q_n'(x_{kn})$ should not increase too quickly as $n \to \infty$. For instance, the assertion of Theorem 11 remains in force if instead of condition $Q_n'(x_{kn}) = 0$ we require

$$Q_n'(x_{kn}) = \frac{o(1)}{\sqrt{1 - x_{kn}^2}} \frac{n^2}{\ln n} \quad \text{as} \quad n \to \infty.$$

The question of the smallest possible increase of the degree of the interpolation polynomial to ensure the convergence of the interpolation process was posed and examined by S. N. Bernstein. For more details about all these questions, see [7] and [11].

**2.5. Operations on interpolation polynomials leading to convergent processes.** Just as we may obtain a convergent process from a divergent series by means of some operation or other, so in the study of interpolation methods we can obtain from interpolation polynomials in definite ways sequences which converge to the interpolated function. The first basic investigation in this direction was carried out by S. N. Bernstein. His results look very simple in the case of the interpolation of periodic functions by trigonometric polynomials (see § 2.2).

From now on by $C_{2\pi}^*$ we shall denote the class of continuous periodic functions of period $2\pi$.

Let the function $f(x)$ belong to $C_{2\pi}^*$. We shall consider its interpolation with simple equidistant points of interpolation

$$x_{kn} = \frac{2k\pi}{2n + 1}, \quad k = 0, 1, \ldots, 2n.$$

Let

$$T_n(x) = a_{0n} + \sum_{k=1}^{n} (a_{kn} \cos kx + b_{kn} \sin kx)$$

be the interpolation polynomial such that

$$f(x_{kn}) = T_n(x_{kn}), \quad k = 0, 1, 2, \ldots, 2n, \quad n = 1, 2, \ldots$$

Let us assume that

$$T_{0n}(x) = a_{0n},$$

$$T_{mn}(x) = a_{0n} + \sum_{k=1}^{m} (a_{kn} \cos kx + b_{kn} \sin kx),$$

$$1 \leqslant m \leqslant n.$$

Suppose finally that

$$U_{sn}(x) = \frac{T_{0n}(x) + T_{1n}(x) + \cdots + T_{sn}(x)}{s+1}.\dagger \qquad (2.22)$$

THEOREM 1 (S. N. Bernstein). *Let $f(x) \in C^*_{2\pi}$. Then any sequence (2.22) of functions $U_{sn}(x)$, $s \leqslant n$, converges uniformly to the function $f(x)$ on the whole axis as $s \to \infty$.*

We now assume that

$$U_n(x) = \frac{T_n\left(x + \frac{\pi}{2n+1}\right) + T_n\left(x - \frac{\pi}{2n+1}\right)}{2}; \qquad (2.23)$$

then the following theorem is true.

THEOREM 2 (S. N. Bernstein). *Let the function $f(x) \in C^*_{2\pi}$. Then the sequence (2.23) of functions $U_n(x)$ converges to the function $f(x)$ uniformly on the whole axis as $n \to \infty$.*

Information on the extensions and further development in this direction may be found in [3], [4], [8], [11] and [12].

**2.6. Generalized interpolation polynomials.** Let the matrix of the points of interpolation (2.21) be given and also some system of functions $\varphi_{kn}(x)$, called *fundamental*

$$\left.\begin{array}{l} \varphi_{11}(x), \\ \varphi_{12}(x), \varphi_{22}(x). \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_{1n}(x), \varphi_{2n}(x), \cdots, \varphi_{nn}(x). \end{array}\right\} \qquad (2.24)$$

Here all the points $x_{kn}$ belong to some given segment $[a, b]$, and all the functions $\varphi_{kn}(x)$ are defined on the same segment.

With the fixed systems (2.21) and (2.24) to each function $f(x)$, defined on the segment $[a, b]$, we may set in correspondence the "polynomial"

$$\Phi_n(x; f) = \sum_{k=1}^{n} f(x_{kn})\varphi_{kn}(x),$$

called a *generalized interpolation polynomial* (with respect to the given system (2.24) of fundamental functions with the given system of points (2.21)).

It is obvious that the Lagrange interpolation polynomial, as formula (2.18) shows, is a generalized interpolation polynomial.

† It is obvious that the method of obtaining the functions $U_s^{(n)}$ is similar in a sense to the Cezaro-Féjer method in the theory of divergent series.

In addition a series of other constructions, met in various sections of mathematics, also leads to the idea of generalized interpolation polynomials. As an example of generalized interpolation polynomials we note the so-called Bernstein polynomials $B_n(x, f)$ (see p. 109). On the assumption that the functions $\varphi_{kn}(x)$ are bounded we may set in correspondence to each pair of the systems (2.21) and (2.24) the functions

$$\lambda_n(x) = \sum_{k=1}^{n} |\varphi_{kn}(x)|, \quad x \in [a, b], \ n = 1, 2, \ldots,$$

and the numbers

$$\lambda_n = \sup_{a \leqslant x \leqslant b} \{\lambda_n(x)\}, \quad n = 1, 2, \ldots$$

Here we have the following theorem (see [9]).

THEOREM. *For any function $f(x) \in C[a, b]$, in order that the sequence of generalized interpolation polynomials $\Phi(x; f)$ (for fixed systems* (2.21) *and* (2.24)) *may converge uniformly on the segment $[a, b]$ to the function $f(x)$, both the following conditions are necessary and sufficient*:

(1) *if $f(x)$ is an algebraic polynomial the sequence $\Phi_n(x; f)$ converges uniformly on the segment $[a, b]$ to the polynomial $f(x)$*;

(2) *there exists $N > 0$ such that for all $n = 1, 2, \ldots$*

$$\lambda_n \leqslant N.$$

Further investigations of the properties of generalized interpolation polynomials can be found in [3], [7], [11].

**2.7. Remarks.** The problems of interpolation which we have considered are related to the class of so-called *problems on the approximation of functions*. As classes $\mathfrak{G}$ of functions, from which we have chosen the functions $g(x)$, which "approximate" the given function $f(x)$ in a defined sense, we have chosen polynomials (algebraic and trigonometric). The "nearness" of the functions $f(x)$ and $g(x)$ was defined simply as the fit of the values of the functions $f(x)$ and $g(x)$ themselves, and sometimes also of some of their derivatives at a finite fixed number of points (the points of interpolation). This is the classical problem of the interpolation of functions. It can be naturally extended to functions of many variables. §§ 2.1, 2.2 and 2.3 have been devoted to the classical problem of interpolation; one of its extensions was given in § 2.6.

As we have been considering only the theory of interpolation in a real domain we have not dwelt on the classical Abel–Goncharov interpolation problem.

This problem consists of the following: given an $n$ times differentiable function $f(x)$ and $n + 1$ points (points of interpolation) $x_0, x_1, \ldots, x_n$ from its domain of definition—to find a polynomial $P(x)$ which satisfies the conditions

$$P^{(k)}(x_k) = f^{(k)}(x_k), \quad k = 0, 1, \ldots, n.$$

The given polynomial $P(x)$ is called an *Abel–Goncharov* interpolation polynomial. As in all interpolation problems, here first of all the question arises of the conditions under which the interpolation polynomials $P(x)$ converge to the function $f(x)$ as $n \to \infty$. It turns out that both from the point of view of the method of solution of this problem and of the classes of functions $f(x)$ for which this problem and its generalization provide interest, it is natural to consider it in the branches of the theory of functions of a complex variable. Moreover the Abel–Goncharov interpolation process turns out to be closely connected with a series of other properties of analytic functions.

The general problem of the interpolation of functions $f(x)$ (one or more variables) of some class $\mathfrak{F}$ by means of functions of class $\mathfrak{G}$ may be formulated as follows: given a finite (sometimes denumerable) number of linear functionals $U_k$, $k = 1, 2, \ldots$, defined both for functions of class $\mathfrak{F}$ and class $\mathfrak{G}$—to find, for a given function $f_0(x) \in \mathfrak{F}$, a function $g_0(x) \in \mathfrak{G}$ such that

$$U_k[g_0] = U_k[f_0], \quad k = 1, 2, \ldots$$

Here, finally great interest is provided by the evaluation of the difference $f_0(x) - g_0(x)$ for arbitrary values of the argument $x$; §§ 2.4 and 2.5 dealt with this question.

Sometimes the determination of "nearness" of the functions $f(x) \in \mathfrak{F}$ and $g(x) \in \mathfrak{G}$ cannot be expressed as an equation of a finite or denumerable number of functionals, as all the values of the argument from the domain of definition of the functions $f(x)$ and $g(x)$ participate in it. The question also occurs of finding a function $g(x) \in \mathfrak{G}$ which only approximates the function $f(x) \in \mathfrak{F}$ with some degree of accuracy or other. Here we are concerned with the typical problem of properly approximating a function, which will be considered in the following paragraphs.

## § 3. UNIFORM APPROXIMATIONS OF FUNCTIONS OF ONE VARIABLE BY POLYNOMIALS AND THEIR GENERALIZATIONS

**3.1. General remarks on the approximation of functions.** It is interesting to study the problem of the approximation of functions also for integral classes of functions $f(x)$. This means that the function $f(x)$ which is being considered is described by some of its own general properties; i.e. it is considered only as representative of some functional class $\mathfrak{F}$. From now on as such an initial class we shall take the class of continuous periodic functions with one and the same period, classes of functions which are summable in some degree,etc. As classes $\mathfrak{G}$ of functions $g(x)$, "approximating" in the defined sense functions of class $\mathfrak{F}$, we shall take for the most part polynomials (algebraic and trigonometric) or integral functions of the exponential type.

In solving the question of the approximation of a function $f(x)$ by functions of class $\mathfrak{G}$ it may happen that either there exist functions $g(x)$ of class $\mathfrak{G}$ as near as we please in our sense to the given function $f(x)$, or this nearness is bounded below by some positive number. Both these cases are met in the study of approximations for the simplest classes of functions and, as we shall see below, lead naturally to quite different future statements of the problems.

As a criterion of nearness in questions of approximation of functions a very important part is played by so-called uniform nearness and mean degree nearness.

If the functions $f(x)$ and $g(x)$ of classes $\mathfrak{F}$ and $\mathfrak{G}$ respectively are defined on some set $E$ then their *uniform nearness* is defined by the quantity

$$\sup_{x \in E} \text{vrai } |f(x) - g(x)|; \dagger$$

in the case where $f(x)$ and $g(x)$ are continuous and $E$ is a bounded closed set, and this quantity can be written simply as

$$\max_{x \in E} |f(x) - g(x)|.$$

† Let a function $\varphi(x)$ be given on some set $M$. For any number $y$ we assume that $M_y = \{x \mid \varphi(x) \geqslant y, x \in M\}$, and that $M^* = \{y \mid \text{mes } M_y = 0\}$. Then by definition

$$\sup_{x \in M} \text{vrai } \varphi(x) = \inf M^*.$$

*Mean degree nearness* with a fixed exponent $p > 0$ is given by the quantity $\int |f(x) - g(x)|^p \, dE$ (here it is assumed that the upper bound $\sup_{x \in E} \text{vrai} \, |f(x) - g(x)|$ and the last integral exist and consequently are finite).

**3.2. Weierstrass' theorem.** Fundamental results which principally solve the question of the approximation of functions, which are continuous on a segment, by polynomials, were put forward by K. Weierstrass. His first theorem is as follows:

THEOREM 1 (K. Weierstrass). *Suppose that on some segment* $[a, b]$ *a continuous function* $f(x)$ *is given. For any* $\varepsilon > 0$ *there exists a polynomial* $P(x)$ *such that for all points* $x \in [a, b]$

$$|f(x) - P(x)| < \varepsilon.$$

This theorem may be paraphrased as follows.

THEOREM 1'. *For any function which is continuous on a segment there exists a sequence of polynomials which converges uniformly to it on this segment.*

Note that the converse statement is also true: any function which is the limit of a uniformly convergent sequence of polynominals on a segment is continuous also on this segment. Thus the possibility of as accurate an approximation of the function as we please on a segment by means of polynomials is a property equivalent to the continuity of the function on this segment.†

THEOREM 2 (K. Weierstrass). *Let* $f(x)$ *be a continuous function of period* $2\pi$; *then for any* $\varepsilon > 0$ *there exists a trigonometric polynomial* $T(x)$ *such that for any real* $x$

$$|f(x) - T(x)| < \varepsilon.$$

This theorem may be given in the following form:

THEOREM 2'. *For any continuous function of period* $2\pi$ *there exists a sequence of trigonometric polynomials which converges uniformly to it on the whole real axis.*

As in the non-periodic case considered above, the possibility of the approximation of a function, which is as accurate as we please, on the whole real axis by means of trigonometric poly-

† The statement of Weierstrass' Theorem 1 is contained in particular in Féjer's Theorem 11 (§ 2.4), but Weierstrass' theorem was proved much earlier than that of Féjer.

nomials is a characteristic property of continuous periodic functions of period $2\pi$.

**3.3. Bernstein polynomials.** The theorems of Weierstrass establish the principal possibility of as accurate an approximation as we please of continuous functions by polynomials. No less important is the question of effectively seeking polynomials which approximate a given continuous function with previously assigned accuracy. Such polynomials are, for instance, those of Féjer (see Theorem 11 in § 2.4) and Bernstein.

DEFINITION. Let a function $f(x)$ be defined on the segment $[0, 1]$. Then its *Bernstein polynomial* of degree $n$ is the polynomial

$$B_n(x, f) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k x^k (1 - x)^{n-k}.$$

It turns out that Bernstein polynomials also effect in particular the uniform approximation of continuous functions, and in the case of sufficiently smooth functions also the uniform approximation of their derivatives.

THEOREM 1. *Let* $f(x) \in C^m$ $[0, 1]$ $(m = 0, 1, 2, \ldots)$; *then the sequences* $\dfrac{d^s B_n(x; f)}{dx^s}$ $(s = 0, 1, 2, \ldots, m\dagger)$ *converge as* $n \to \infty$ *to the function* $\dfrac{d^s f(x)}{dx^s}$ *uniformly on the segment* $[0, 1]$.

In the case where $m = 0$ this theorem has been proved by S. N. Bernstein, and for the case of an arbitrary $m$ $(m = 0, 1, 2, \ldots)$ generalized by I. N. Khlodovskii.

From this theorem with $m = 0$ follows once again Weierstrass' Theorem 1 (§ 3.2) and in addition there is clearly indicated a polynomial which approximates a given function.

THEOREM 2 (I. Natanson). *Let the function* $f(x)$ *be bounded on the segment* $[0, 1]$ *and at the point* $x_0 \in [0, 1]$ *have a derivative, then*

$$\lim_{n\to\infty} B_n'(x_0; f) = f'(x_0).$$

Theorem 1 shows that (with $m = 0$) any continuous function may be presented as the limit of the uniformly convergent sequence of its Bernstein polynomials; thus it proves to be possible to evaluate the speed of its convergence by the modulus of continuity $\omega(x; f)$ of the function $f(x)$ (see § 2.4).

† By the zero derivative of the function $(m = 0)$ we mean as usual the function itself.

THEOREM 3 (T. Popoviciu). *Let* $f(x) \in C[a, b]$, *then*

$$|B_n(x; f) - f(x)| \leqslant \frac{3}{2}\,\omega\left(\frac{1}{\sqrt{n}}; f\right).$$

For a function which is twice differentiable at some point we have the following evaluation:

THEOREM 4 (E. Voronovskaya). *Suppose the function* $f(x)$ *is bounded on the segment* $[0, 1]$ *and that at the point* $x_0 \in [0, 1]$ *the derivative* $f''(x_0)$ *exists; then as* $n \to \infty$

$$B_n(x_0; f) - f(x_0) = \frac{f''(x_0)}{2n}\,x_0(1 - x_0) + \frac{o(1)}{n}.$$

Other properties of the Bernstein polynomials which we have been considering and their generalization can be found in [3], [8], [11] and [12].

**3.4. The best uniform approximations of functions by polynomials of a given degree.** Bernstein's first theorem (see § 3.2) shows that any function which is continuous on a segment with any preassigned degree of accuracy may be approximated by a polynomial. Here we have neither considered how the accuracy is changed as the degree of the polynomial increases, nor the question of what best accuracy of approximation may be guaranteed by confining ourselves to polynomials of a fixed degree, nor finally the question of how to find among all the polynomials of given degree which of them deviates least from the given function, or even the question of its existence.

The problem of polynomials which deviate least from the given function was first put forward and investigated more than a century ago, in 1853, by the great Russian mathematician P. L. Chebyshev,† who rightfully is also considered to be the founder of the theory of the approximation of functions.

THEOREM (E. Borel). *Let* $f(x) \in C[a, b]$. *Then for any* $n = 0, 1, 2, \cdots$ *there exists a polynomial* $P(x) \in \mathfrak{P}_n$ *such that*

$$\max_{a \leqslant x \leqslant b} |P(x) - f(x)| = E_{\mathfrak{P}_n}(f)\ddagger \tag{3.1}$$

† Chebyshev, P. L., "The theory of mechanisms known by the name of parallelograms, 1853, *Complete collection of works*, vol. II, M-L, 1947.

‡ For a definition of $E_{\mathfrak{P}_n}(f)$ see p. 98.

The polynomial $P(x)$ given in Borel's theorem, i.e. possessing the property (3.1). is called the *best approximating polynomial of* the function $f(x)$. We can show that in class $\mathfrak{P}_n$ the best approximating polynomial of the function $f(x)$ is unique.

We shall determine the best approximation of periodic functions by means of trigonometric polynomials of given degree. If $\mathfrak{T}_n$, $n = 0, 1, \ldots$, is the aggregate of trigonometric polynomials of degree not higher than $n$ (see § 2.2) then for each periodic function $f(x)$ of period $2\pi$ the value

$$E_{\mathfrak{T}_n}(f) = \inf_{T(x) \in \mathfrak{T}_n} \sup_{-\infty < x < \infty} |T(x) - f(x)|$$

is called the *best approximation of the function $f(x)$ by trigonometric polynomials* of degree not higher than $n$. As in the non-periodic case, with any fixed $n = 0, 1, \ldots$, for any function $f(x) \in C^*_{2\pi}$ there always exists a trigonometric polynomial (and that unique) $T(x) \in \mathfrak{T}_n$ such that

$$\max_{-\infty < x < \infty} |T(x) - f(x)| = E_{\mathfrak{T}_n}(f).$$

This polynomial is called the *best approximating trigonometric polynomial* of the function $f(x)$ in the class $\mathfrak{T}_n$.

The best approximations, defined in the present section, are said to be *uniform* best approximations as distinct from other best approximations, which will be considered below (see § 5).

If $f(x) \in C[a, b]$, then according to Weierstrass' first theorem (see § 3.2)

$$\lim_{n \to \infty} E_{\mathfrak{P}_n}(f) = 0.$$

Similarly in the case where $f(x) \in C^*_{2\pi}$, from Weierstrass' second theorem (see § 3.2) it follows that

$$\lim_{n \to \infty} E_{\mathfrak{T}_n}(f) = 0.$$

The best approximations $E_{\mathfrak{P}_n}$ and $E_{\mathfrak{T}_n}(f)$ obviously do not increase as $n$ increases, $n = 0, 1, 2, \ldots$, i.e.

$$E_{\mathfrak{P}_0}(f) \geqslant E_{\mathfrak{P}_1}(f) \geqslant \cdots \geqslant E_{\mathfrak{P}_n}(f) \geqslant \cdots,$$
$$E_{\mathfrak{T}_0}(f) \geqslant E_{\mathfrak{T}_1}(f) \geqslant \cdots \geqslant E_{\mathfrak{T}_n}(f) \geqslant \cdots$$

It was shown by S. N. Bernstein that the converse is also true: whatever the sequence of numbers

$$\alpha_0 \geqslant \alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant \cdots \geqslant 0, \qquad \lim_{n\to\infty} \alpha_n = 0, \tag{3.2}$$

and whatever the segment $[a, b]$, there exists a function $f(x) \in C[a, b]$ such that for all $n = 0, 1, 2, \ldots$

$$E_{\mathfrak{P}_n}(f) = \alpha_n.$$

In addition Bernstein showed that for any sequence (3.2) there exists a function $f(x) \in C^*_{2\pi}$ such that for all $n = 0, 1, 2, \ldots$

$$E_{\mathfrak{T}_n}(f) = \alpha_n.$$

**3.5. Chebyshev's results for best approximations.** Chebyshev pointed out the following necessary and sufficient condition for the given polynomial to be the best approximating polynomial of a continuous function.

THEOREM 1 (P. L. Chebyshev). *Let $f(x) \in C[a, b]$, $n$ be a fixed non-negative whole number, $P(x)$ a polynomial: $P(x) \in \mathfrak{P}_n$ (see § 3.4) and*

$$A = \max_{a\leqslant x\leqslant b} |f(x) - P(x)|.$$

*The polynomial $P(x)$ is the best approximating polynomial of the function $f(x)$ in the class $\mathfrak{P}_n$ when and only when there exist, on the segment $[a, b]$, $n + 2$ points $x_i$, $i = 1, 2, \ldots, n + 2$,*

$$a \leqslant x_1 < x_2 < \cdots < x_{n+2} \leqslant b,$$

*such that*

$$|Q(x_i) - f(x_i)| = A, \quad i = 1, 2, \ldots, n + 2,$$

*and the sign of the difference $Q(x_i) - f(x_i)$ is changed with passage from each point $x_i$ to the following $x_{i+1}$ ($i = 1, 2, \ldots, n + 1$).*

In virtue of the definition given above, for the conditions of the theorem to be fulfilled we must obviously have

$$A = E_{\mathfrak{P}_n}(f).$$

A similar statement holds also for the case of best approximations by trigonometric polynomials.

THEOREM 2 (P. L. Chebyshev). *Let $f(x) \in C^*_{2\pi}$, $n$ be a fixed non-negative whole number and $T(x)$ a trigonometric polynomial, $T(x) \in \mathfrak{T}_n$* (see § 3.4). *In addition let*

$$A = \max_{-\infty < x < \infty} |f(x) - T(x)|.$$

*The polynomial $T(x)$ is the best approximating trigonometric polynomial of the function $f(x)$ in the class $\mathfrak{T}_n$ when and only when there exist $2n + 2$ points $x_i$, $i = 1, 2, \ldots, 2n + 2$,*

$$0 \leqslant x_1 < x_2 < \cdots < x_{2n+2} < 2\pi,$$

*on the interval $[0, 2\pi]$ such that*

$$|T(x_i) - f(x_i)| = A, \quad i = 1, 2, \ldots, 2n + 2,$$

*and the sign of the difference $T(x_i) - f(x_i)$ changes with passage from each point $x_i$ to the following $x_{i+1}$ $i = 1, 2, \ldots, 2n + 1$).*

It is obvious that with the fulfilment of the conditions of the theorem

$$A = E_{\mathfrak{T}_n}(f).$$

In Theorems 1 and 2 we have naturally assumed that the function $f(x)$ is not a polynomial (algebraic or trigonometric respectively) of degree not above $n$, because in this case the function $f(x)$ simply coincides with its own best approximating polynomial of degree not higher than $n$.

As an example we will give a problem on the search in class $\mathfrak{P}_{n-1}$ for a polynomial $P(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$, which deviates least from the function $f(x) = x^n$ on the segment $[-1, 1]$. To put it otherwise, if

$$r(x) = x^n - (a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}), \tag{3.3}$$

it is required to choose the coefficients $a_0, a_1, \ldots, a_{n-1}$, so that the quantity $\max_{-1 \leqslant x \leqslant 1} |r(x)|$ has the smallest possible value. Obviously we may write in the form (3.3) any polynomial of degree $n$ with coefficient 1 for the leading term. Thus the presentation of the problem is equivalent to the problem of finding, in the class of all polynomials of degree $n$ with the coefficient 1 for the leading term, a polynomial with the least absolute value on the segment $[-1, 1]$ or, what is the same, a polynomial of the given class which deviates least from zero on the segment $[-1, 1]$.

Such polynomials were discovered by P. L. Chebyshev and are called *Chebyshev polynomials* (see § 2.1).

**3.6. The approximate formation of best approximating Chebyshev polynomials.** In the present section we shall consider some methods of the approximate calculation of best approximating Chebyshev polynomials for functions which are continuous on a segment. The following theorem gives the value from below for the best approximation of the given function by means of a polynomial chosen in a definite way.

THEOREM 1 (de la Vallée Poussin). *Let* $f(x) \in C[a, b]$, *and* $n + 2$ *points*

$$x_0 < x_1 < x_2 < \cdots < x_{n+1};$$

*be given on the segment* $[a, b]$. *Also suppose that for some polynomial*

$$Q(x) \in \mathfrak{P}_n$$

*one of the two following systems of inequalites holds*:

$$\begin{aligned} f(x_0) - Q(x_0) &> 0, \\ f(x_1) - Q(x_1) &< 0, \\ f(x_2) - Q(x_2) &> 0 \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

*or*

$$\begin{aligned} f(x_0) - Q(x_0) &< 0, \\ f(x_1) - Q(x_1) &> 0, \\ f(x_2) - Q(x_2) &< 0 \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

*Then for the best approximation* $E_{\mathfrak{P}_n}(f)$

$$E_{\mathfrak{P}_n}(f) \geqslant \min_{i=0, 1, \ldots, n} |f(x_i) - Q(x_i)|.$$

As we already know (see § 3.5) the equality is attained here when and only when $Q(x)$ is the best approximating polynomial.

Approximate methods of finding the best approximating polynomial are based first of all on the following theorem.

THEOREM 2. *Let* $f(x) \in C[a, b]$, $P_0(x) \in \mathfrak{P}_n(x)$, *where* $P_0(x)$ *is the best approximating polynomial of the function* $f(x)$ *in* $\mathfrak{P}_n$, *i.e.*

$$\begin{aligned} E_{\mathfrak{P}_n}(f) &= \max_{a \leqslant x \leqslant b} |f(x) - P_0(x)| = \\ &= \min_{P(x) \in \mathfrak{P}_n} \max_{a \leqslant x \leqslant b} |f(x) - P(x)|. \end{aligned}$$

*Then for any* $\varepsilon > 0$ *there exists* $\delta = \delta(\varepsilon)$ *such that as soon as for the polynomial* $P(x) \in \mathfrak{P}_n$ *the following inequality is satisfied*

$$|f(x) - P(x)| < E_{\mathfrak{P}_n}(f) + \delta, \quad a \leqslant x \leqslant b,$$

*then*

$$|P(x) - P_0(x)| < \varepsilon.$$

It is not difficult to see that as $\varepsilon$ approaches zero the coefficients of the given polynomials $P(x)$ approach the coefficients of the corresponding powers of $x$ in the polynomial $P_0(x)$.

Theorem 2 shows that, in order to find the best approximating polynomial of a given function $f(x)$ with a defined degree of accuracy, it is sufficient to find a polynomial whose deviation from the function $f(x)$ differs sufficiently little from its best approximation.

We now consider one of the algorithms for calculating the approximations of the Chebyshev best approximating polynomials.

Let $f(x) \in C[a, b]$ and $P(x)$ be a polynomial of degree not higher than $n$: $P(x) \in \mathfrak{P}_n$. We form the difference

$$r(x) = f(x) - P(x),$$

and assume that there exist at least $n + 2$ points

$$x_0 < x_1 < \cdots < x_{n+1}, \tag{3.4}$$

at which this difference has successively alternating signs:

$$\operatorname{sign} r(x_0) = -\operatorname{sign} r(x_1) = \cdots = (-1)^{n+1} \operatorname{sign} r(x_{n+1}). \tag{3.5}$$

Now let

$$L = L(P) = \max_{a \leqslant x \leqslant b} |r(x)| \tag{3.6}$$

and

$$A(x_0, x_1, \ldots, x_{n+1}; P) = \min_{i=0, 1, \ldots, n+1} |r(x_i)|.$$

We now choose a system of points $\{x_i^{(0)}\}_{i=0}^{n+1}$ such that conditions (3.5) are satisfied and that the quantity

$$A(x_0, x_1, \ldots, x_{n+1}; P) \text{ with } x_i = x_i^{(0)}, i = 0, 1, 2, \ldots, n,$$

has the largest possible value, i.e.

$$A(P) = A(x_0^{(0)}, x_1^{(0)}, \ldots, x_n^{(0)}; P) = \max_{x_0 < x_1 < \ldots < x_{n+1}} A(x_0, x_1, \ldots, x_{n+1}; P), \tag{3.7}$$

and also

$$\max_{i=0,1,\dots,n+1} |r(x_i^{(0)})| = L. \tag{3.8}$$

For the fulfilment of (3.7) and (3.8) it is sufficient to take as points $\{x_i^{(0)}\}_{i=0}^{n+1}$ the respective successive extreme points of the function $r(x)$.

From Theorem 1 it follows that

$$E_{\mathfrak{P}_n}(f) \geqslant A(P).$$

We now give an algorithm with whose help we can form a "complement" to the polynomial $P(x)$, in the form of some polynomial $Q_1(x) \in \mathfrak{P}_n$, such that the polynomial $P_1(x) = P(x) + Q_1(x)$ differs even less from the best approximating polynomial than the required polynomial $P(x)$.

We shall define the polynomial $Q_1(x)$ as the polynomial which best approximates the function $r(x)$ on the set of $n + 2$ points $x_0^{(0)}, x_1^{(0)}, \dots, x_{n+1}^{(0)}$, i.e. as the polynomial for which

$$\begin{aligned} \varrho &= \max_{i=0,1,\dots,n+1} |r(x_i) - Q_1(x_i)| = \\ &= \inf_{Q(x) \in \mathfrak{P}_n} \max_{i=0,1,\dots,n+1} |r(x_i) - Q(x_i)|. \end{aligned}$$

The polynomial $Q_1(x)$ (see § 7.3) satisfies the following system of $n + 1$ equations:

$$\begin{aligned} r(x_0) - Q_1(x_0) &= -[r(x_1) - Q_1(x_1)] = \dots = \\ &= (-1)^{n+1} [r(x_{n+1}) - Q_1(x_{n+1})] = \varrho \operatorname{sign} r(x_0). \end{aligned} \tag{3.9}$$

If $Q_1(x) = c_0 + c_1 x + \dots + c_n x^n$ the system (3.9) may be rewritten as a system of $n + 2$ linear equations in $c_0, c_1, \dots, c_n$ and $\varrho$:

$$\left.\begin{aligned} \varrho \operatorname{sign} r(x_0) + c_0 + c_1 x_0 + \dots + c_n x_0^n &= r(x_0), \\ -\varrho \operatorname{sign} r(x_1) + c_0 + c_1 x_1 + \dots + c_n x_1^n &= r(x_1), \\ (-1)^{n+1} \operatorname{sign} r(x_{n+1}) + c_0 + c_1 x_{n+1} + \dots + c_n x_{n+1}^n &= \\ = r(x_{n+1}). & \end{aligned}\right\} \tag{3.10}$$

Assuming that

$$d_i = \frac{1}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_{i+1} - x_i) \cdots (x_{n+1} - x_i)},$$

from (3.10) we obtain

$$\varrho = \frac{d_0 |r(x_0)| + d_1 |r(x_1)| + \dots + d_{n+1} |r(x_{n+1})|}{d_0 + d_1 + \dots + d_{n-1}}.$$

Consequently we now know the values of the polynomial $Q_1(x)$ at the points $x_i^{(0)}$:

$$Q_1(x_i^{(0)}) = r(x_i^0) - \varrho \operatorname{sign} r(x_i^0), \quad i = 0, 1, \ldots, n+1,$$

and therefore the polynomial $Q_1(x)$ itself may be directly written down from the formula for the Lagrange interpolation polynomial (see § 2.1) using any $n+1$ points taken from the system of points $\{x_i^{(0)}\}_{i=0}^{n+1}$.

Let us assume that

$$P_1(x) = P(x) + Q_1(x)$$

and (see (3.7))

$$A_1 = A(P_1);$$

then there exists a number $\theta$, $0 < \theta < 1$ such that

$$E_{\mathfrak{P}_n}(f) - A_1 < \theta[E_{\mathfrak{P}_n}(f) - A].$$

If we proceed from the polynomial $P_1(x)$ and form a polynomial $P_2(x)$, in the same way as the polynomial $P_1(x)$ was formed from the polynomial $P(x)$, etc., we shall obtain a sequence of polynomials $P_k(x) \in \mathfrak{P}_n$, $k = 1, 2, \ldots$ for which, on the assumptions (see (3.6), (3.7) and (3.8)) that

$$A_k = A(P_k) \quad \text{and} \quad L_k = L(P_k),$$

we shall have

$$E_{\mathfrak{P}_n}(f) - A_k < \theta^k (E_{\mathfrak{P}_n}(f) - A),$$

$$L_k - E_{\mathfrak{P}_n}(f) < L_k - A_k < \frac{E_{\mathfrak{P}_n}(f) - A}{1 - \theta},$$

where $0 < \theta < 1$ and $\theta$ does not depend on $k$ ($\theta$ is determined by the function $f(x)$ and the original polynomial $P(x)$).

Hence it follows that

$$\lim_{k \to \infty} A_k = \lim_{k \to \infty} L_k = E_{\mathfrak{P}_n}(f)$$

and then (see Theorem 2) the sequence of polynomials $P_k(x)$ converges uniformly on the segment $[a, b]$ to the best Chebyshev approximating polynomial of the function $f(x)$.

We must note that there also exist other methods of approximate formation of best approximating polynomials. One of them is based as in the above method, on the choice of some correction which improves the approximation chosen *a priori*; others on completely different ideas connected for instance with the approximation of

functions in the mean. All these questions are clarified in detail in the monographs [15], [18] and [20].

**3.7. The best approximations to continuous and differentiable periodic functions by trigonometric polynomials of a given degree.** We now turn to the study of the influence of properties of the continuous function $f(x)$ on the speed at which its best approximations $E_{\mathfrak{T}_n}(f)$ or, in the periodic case, $E_{\mathfrak{T}_n}(f)$ approach zero. We begin with the periodic case. Let $C^*_{2\pi}$, as before, be the class of continuous periodic functions of period $2\pi$.

THEOREM 1 (D. Jackson). *Let the function $f(x) \in C^*_{2\pi}$ have $k$ continuous derivatives and let $\omega(\delta; f^{(k)})$ be the modulus of continuity of the $k$th derivative $f^{(k)}(x)$. Then*

$$E_{\mathfrak{T}_n}(f) \leqslant \frac{c\omega\left(\frac{1}{n}; f^{(k)}\right)}{n^k},$$

*where $c$ is a constant which depends only on $k$.*

For the proofs of similar theorems integral presentations of the form

$$T(x) = \int_{-\pi}^{\pi} K_n(t - x)\, f(t)\, dt$$

have a fundamental significance for trigonometric polynomials $T(x) \in \mathfrak{T}_n$, which approximate sufficiently accurately the given function $f(x)$ (see also § 4.2).

Similar constructions offer independent interest. As an example we may note one of the results of this type (for which, in particular, Theorem 1 follows directly if $k = 0$).

Let $f(x) \in C^*_{2\pi}$, and $\omega(\delta; f)$ be the modulus of continuity of the function $f(x)$. Then the function

$$T(x) = \frac{3}{2\pi n(2n^2 + 1)} \int_{-\pi}^{\pi} \left[ \frac{\sin \frac{n(t - x)}{2}}{\sin \frac{t - x}{2}} \right]^4 f(t)\, dt$$

is a trigonometric polynomial of class $\mathfrak{T}_{2n-2}$, for which

$$|T(x) - f(x)| \leqslant 6\omega\left(\frac{1}{n}; f\right).$$

The connection between the best approximations $E_{\mathfrak{T}_n}(f)$ and the moduli of continuity of the function $f(x)$ has been explained very fully in the works of S. B. Stechkin, in which attention is given to moduli of continuity of higher orders. We shall formulate a definition of the latter for any periodic function $f(x)$ of period $2\pi$: the quantity

$$\omega_p(\delta; f) = \sup_{|h| \leq \delta} \sup_{-\infty < x < \infty} \left| \sum_{k=0}^{p} (-1)^{p-k} C_p^k f(x + kh) \right|$$

is called the *modulus of continuity* of order $p = 1, 2, 3, \ldots$ of the function $f(x)$. Obviously

$$\omega_1(\delta; f) = \omega(\delta; f).$$

Note also that if $f(x) \in C_{2\pi}^*$ and $|f(x)| \leq M$ for all $x$, then for the best approximations of the function $f(x)$ we have the inequality established by J. Favard

$$E_{\mathfrak{T}_n}(f) \leq \frac{MK_r}{(n+1)^r}, \quad n = 0, 1, 2, \ldots,$$

where

$$K_r = \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu(r+1)}}{(2\nu+1)^{r+1}}, \quad r = 1, 2, 3, \ldots$$

THEOREM 2 (S. B. Stechkin). *Let $f(x) \in C_{2\pi}^*$ and let the derivative $f^{(r)}(x) \in C_{2\pi}^*$ ($r = 1, 2, 3, \ldots$) exist; then*

$$E_{\mathfrak{T}_n}(f) \leq \frac{2K_r}{(n+1)^r} E_{\mathfrak{T}_n}(f^{(r)}).$$

THEOREM 3. *Let $f(x) \in C_{2\pi}^*$; then*

$$E_{\mathfrak{T}_n}(f) \leq c_p \omega_p\left(\frac{1}{n}; f\right), \quad p = 1, 2, 3, \ldots, \; n = 0, 1, 2, \ldots,$$

*where the constant $c_p$ depends on $p$.*

For $p = 1$ the theorem was proved by Jackson (see Theorem 1) and for $p > 1$ by Stechkin.

We may also, conversely, evaluate the modulus of continuity of the function $f(x)$ by means of its best approximations.

THEOREM 4 (S. B. Stechkin). *Let $f(x) \in C_{2\pi}^*$; then*

$$\omega_p(\delta; f) \leq \frac{d_p}{n^p} \sum_{\nu=1}^{n} \nu^{p-1} E_{\mathfrak{T}_n}(f), \quad p = 1, 2, \ldots,$$

*where $0 < \delta \leq 1$, $n \leq 1/\delta$, and the constant $d$ depends on $k$.*†

† A similar result was obtained earlier by A. F. Timan in the case of space $L_p$, $1 < p < \infty$.

We shall now dwell in more detail on the particular case where the order of increase of the value of $E_{\mathfrak{T}_n}(f)$ has an exponential character, more accurately where $E_{\mathfrak{T}_n}(f)$ has the order $1/n^r$ and $r > 0$. For this we shall introduce, for convenience, the idea of $H$-classes.

Let $f(x) \in C^*_{2\pi}$, and let $r > 0$, $r = \bar{r} + \alpha$, where $\bar{r}$ is a non-negative whole number, $0 < \alpha \leqslant 1$. Further let the function $f(x)$ have a continuous derivative of order $\bar{r}$.† The function $f(x)$ is called a *function of class* $H^{(r)}_*(M)$ if there exists a constant $M > 0$ such that for any real $x$ and $h$

$$|f^{(\bar{r})}(x + h) - 2f^{(\bar{r})}(x) - f^{(\bar{r})}(x - h)| \leqslant M|h|^\alpha. \qquad (3.11)$$

Note that in the case where $0 < \alpha < 1$ the condition (3.11) is equivalent to the usual Hölder condition of degree $\alpha$ for the $\bar{r}$th derivative:

$$|f^{(\bar{r})}(x + h) - f^{(\bar{r})}(x)| \leqslant K|h|^\alpha.$$

The union of the classes $H^{(r)}_*(M)$ for all possible constants $M > 0$ with fixed $r$ is represented by $H^{(r)}_*$:

$$H^{(r)}_* = \bigcup_{M>0} H^{(r)}_*(M).$$

Finally we shall represent by $\mathrm{Lip}^*_M 1$ the aggregate of functions $f(x) \in C^*_{2\pi}$ which satisfy the Lipschitz condition

$$|f(x + h) - f(x)| < M|h|,$$

and assume that

$$\mathrm{Lip}^* 1 = \bigcup_M \mathrm{Lip}^*_M 1.$$

It is obvious that with $0 < \alpha < 1$ the following inclusion is true:

$$\mathrm{Lip}^* 1 \subset H^{(1)}_* \supset H^{(\alpha)}_*. \qquad (3.12)$$

THEOREM 5. *Let* $F(x) \in H^{(r)}_*(M)$, $r > 0$, $r = \bar{r} + \alpha$, $\bar{r}$ *is a non-negative whole number,* $0 < \alpha \leqslant 1$; *then*

$$E_{\mathfrak{T}_n}(f) \leqslant \frac{cM}{n^r},$$

*where the constant* $c$ *does not depend either on* $M$ *or* $n$.

† In our further work a sufficient requirement is that the function $f(x)$ should have an absolutely continuous derivative of order $\bar{r} - 1$.

With $\alpha < 1$ this theorem (which is a corollary of Theorem 1) was proved by D. Jackson, and with $\alpha = 1$ by A. Zygmund.

The theorem which is the converse of Theorem 5 in the case where $0 < \alpha < 1$ is S. N. Bernstein's, revised by de la Vallée Poussin. For its proof the inequality which is called *Bernstein's inequality* has fundamental significance. It connects the maximum absolute value of the derivative of the trigonometric polynomial with the maximum absolute value of the polynomial itself: let $T(x)$ be a trigonometric polynomial of degree $n$, then

$$\max_{-\infty<x<\infty} |T'(x)| \leqslant n \max_{-\infty<x<\infty} |T(x)|.$$

This inequality finds a varied application also in a number of other questions in the theory of functions.

THEOREM 6. *Let $f(x) \in C^*_{2\pi}$, $r > 0$ and $r = \bar{r} + \alpha$, $\bar{r}$ is a non-negative whole number, $0 < \alpha \leqslant 1$. Then if for all natural numbers $n$* †

$$E_{\mathfrak{T}_n}(f) \leqq \frac{A}{n^r} \tag{3.13}$$

*where $A$ is a constant, $f(x) \in H^{(r)}_*(M)$, where $M \leqslant c[A + \max_{-\infty<x<\infty} |f(x)|]$ the constant $c$ not depending on $A$ and $f$.*

The case where $0 < \alpha < 1$ was, as we have already remarked, considered by S. N. Bernstein and where $\alpha = 1$ by A. Zygmund.

We can see that Theorems 5 and 6 together give the necessary and sufficient condition for the function $f(x)$ to belong to the class $H^{(r)}_*$, expressed in terms of best approximations.

Concerning the case where the function $f(x) \in C^*_{2\pi}$ has a derivative $f^{(k)}(x) \in \mathrm{Lip}^*1$, from Theorem 5 and the inclusion (3.12) it follows that there exists a constant $A > 0$ such that

$$E_{\mathfrak{T}_n}(f) \leqslant \frac{A}{n^{k+1}}. \tag{3.14}$$

From the fact, however, that for some function $f(x) \in f(x^*_{2\pi})$ and all natural numbers $n$ inequality (3.14) holds, generally speaking it does not follow that the function $f(x)$ has a derivative $f^{(k)}(x)$ belonging to class $\mathrm{Lip}^* 1$. For example for the function $f(x) = \sum_{k=1}^{\infty} \frac{\sin kx}{k^2}$ the inequality $E_{\mathfrak{T}_n}(f) < \frac{1}{n}$ holds, but $f(x) \bar{\in} \mathrm{Lip}^* 1$.

† The statement of the theorem remains in force if we require the inequality (3.13) to be satisfied only for some sequence $n$ which forms an increasing geometric sequence.

On the other hand, the value (3.14) for the functions considered cannot be improved in the sense of increasing the exponent of $n$ because this would mean, according to Theorem 6, that the function being considered in fact has also a derivative of order $k + 1$.

Note also that the question of how to describe in terms of best approximations the class which consists of all functions $f(x) \in C^*_{2\pi}$ which have a bounded derivative, and only of such functions, and also the general question of the possibility of this, remains at the present time open.

Let us now consider an infinitely differentiable function.

THEOREM 7 (S. N. Bernstein). *In order that the function $f(x) \in C^*$ should have derivatives of any order it is necessary and sufficient that for any $r > 0$*

$$\lim_{n\to\infty} n^r E_{\mathfrak{T}_n}(f) = 0.$$

To formulate the result which characterizes the best approximations of analytic functions we introduce the following definition: a function $f(x) \in C^*_{2\pi}$ is called *a function of class $A^*_{2\pi}$* if there exists $R > 0$ such that the function $f(x)$, with $|x - x_0| < R$, can be expanded in a power series

$$f(x) = \sum_{k=0}^{\infty} c_k (x - x_0)^k,$$

whose coefficients $c_k$ depend generally speaking on $x_0$; $c_k = c_k(x_0)$. If $R = +\infty$ the function $f(x)$ is called an *integral* function, and the subclass of integral functions of class $A^*_{2\pi}$ is denoted by $\mathfrak{G}^*_{2\pi}$.

THEOREM 8 (S. N. Bernstein). *Let $f(x) \in C^*_{2\pi}$. Then $f(x) \in A^*_{2\pi}$ when and only when*

$$E_{\mathfrak{T}_n}(f) \leqslant kq^n.$$

*where $k$ and $q < 1$ are constants.*

THEOREM 9 (S. N. Bernstein). *Let $f(x) \in C^*_{2\pi}$; then $f(x) \in \mathfrak{G}^*_{2\pi}$ when and only when*

$$\lim_{n\to\infty} \sqrt[n]{E_{\mathfrak{T}_n}(f)} = 0.$$

In a number of cases we can obtain not only the value of the order of the best approximations for some class of functions but also an accurate value of the best approximation. We note one of the results of this type.

Let $r$ be a non-negative integer, $K > 0$. We denote by $W_*^{(r)}(K)$ the aggregate of all periodic functions of period $2\pi$ which have a continuous derivative of order $r$ such that

$$|f^{(r)}(x)| \leqslant K, \quad -\infty < x < \infty.$$

THEOREM 10 (J. Favard, N. I. Akhiezer and M. G. Krein).

$$\sup_{f \in W_*^{(r)}(K)} E_{\mathfrak{T}_{n-1}}(f) = \frac{4K}{\pi n^2} \sum_{k=0}^{\infty} \frac{(-1)^{k(r+1)}}{(2k+1)^{r+1}}.$$

Up to now we have considered basically the direct and inverse relations between the properties of functions of a given class and the speed with which the best approximations of functions of a given class approach zero. We must add that a group of problems of another type is interesting. Suppose we are given some class of continuous periodic functions; the question arises; can we say anything on the whole about the properties of the best approximating polynomials of functions of a given class on the basis of some property or other of the functions of the given class themselves? No less interesting is the converse problem. Note as an example a similar result. For the function $f(x) \in C_{2\pi}^*$ we denote by $T_n^*(x; f)$ the trigonometric best approximating polynomial of degree not higher than $n$: $T_n^*(x; f) \in \mathfrak{T}_n$, and for any $\alpha$, $0 < \alpha < 1$, for simplicity we assume by definition that $\mathrm{Lip}_M^* \alpha = H_*^{(\alpha)}(M)$.

THEOREM 11 (S. B. Stechkin). *In order that the function $f(x)$ may belong to the class $\mathrm{Lip}_M^* \alpha$, $0 < \alpha \leqslant 1$, it is necessary and sufficient that a constant $c > 0$ can be found, not depending on $n$, such that*

$$T_n^*(x; f) \in \mathrm{Lip}_{cM}^* \alpha$$

*for all $n = 1, 2, \ldots$*

Information about other results and further development of the theory of best approximations of functions of class $C_{2\pi}^*$ may be found in [1], [3], [4], [7], [8], [11] and [12].

**3.8. The best approximations of continuous and differentiable functions by algebraic polynomials of a given degree.** In the non-periodic case the question of the evaluation of the order of the best approximations of functions by algebraic polynomials is much more complicated than the similar problem for the periodic case considered in the previous section. Originally, in the works of D. Jackson, S. N. Bernstein, A. Zygmund and P. Montel, a series

of problems on the smoothness characteristic of a function in terms of best approximations was investigated, not for the whole segment $[a, b]$, on which the function is given, but only for some inner segment $[a', b']$ lying in the interval $(a, b)$. Only in very recent times, thanks to the results obtained by S. M. Nikol'skii, A. F. Timan and V. K. Dzyadyk, has success been obtained in a sense in getting a complete solution of a definite group of questions, this being done not in terms of best approximations $E_{\mathfrak{P}_n}(f)$ but by an idea sufficiently near to this. Here we first of all formulate the results. We begin with an extension of D. Jackson's theorem, due to A. F. Timan.

THEOREM 1 (A. F. Timan). *Let a number $k$ $(k = 0, 1, 2, \ldots)$ and a function $f(x) \in C^{(k)}[a, b]$ be given. Then for any $n = 0, 1, 2, \ldots$ there exists a polynomial $P(x) \in \mathfrak{P}_n$ such that for any $x \in [a, b]$*

$$|f(x) - P(x)| \leqslant \frac{c}{n^k}\left[\sqrt{(b-x)(x-a)} + \frac{\left|x - \frac{a+b}{2}\right|}{n}\right]^k \times$$

$$\times\, \omega\left[\frac{1}{n}\left(\sqrt{(b-x)(x-a)} + \frac{\left|x - \frac{a+b}{2}\right|}{n}\right),\ f^{(k)}\right],$$

*where the constant $c$ does not depend on $n$ or $x$.*

We see that, as distinct from Jackson's theorem for the periodic case (see § 3.7),† the given evaluation of the approximation of the function by a polynomial depends on the position of the point $x$ on the segment. Only evaluations of such a kind permit us in the non-periodic case to reverse Theorem 1 for some classes of functions (see Theorem 3 below). A result similar to Theorem 1, which was qualitatively new for his time, was first obtained by S. M. Nikol'skii.

For the formulation of further results, on an analogy with classes $H^{(r)}_*$, we shall define classes $H^{(r)}$ for the non-periodic case.

Let $r > 0$, $r = \bar{r} + \alpha$, $\bar{r}$ is a non-negative whole number, $0 < \alpha \leqslant 1$. The function $f(x)$ is called *a function of class $H^{(r)}$* on the segment $[a, b]$, if $f(x) \in C^{(\bar{r})}[a, b]$ and there exists a constant $M > 0$ such that in the case where $0 < \alpha < 1$

$$|f^{(\bar{r})}(x + h) - f^{(\bar{r})}(x)| \leqslant M|h|^{\alpha},$$

† And also as distinct from Jackson's corresponding theorem for the non-periodic case which is not cited here.

and in the case where $\alpha = 1$

$$|f^{(r)}(x + h) - 2f(x) + f(x - h)| \leqslant M|h|,$$

where $x \in [a, b]$, $x \pm h \in [a, b]$.

THEOREM 2 (A. F. Timan and V. K. Dzyadyk). *Let the function $f(x)$ be a function of class $H^{(r)}$ on the segment $[a, b]$. Then for any $n = 0, 1, \ldots$ there exists a polynomial $P_n(x) \in \mathfrak{P}_n$ such that*

$$|f(x) - P_n(x)| \leqslant \frac{c}{n^r}\left\{[\sqrt{(x - a)(b - x)}]^r + \frac{1}{n^r}\right\},$$

*where the constant $c$ does not depend on $n$ or $x$.*†

The converse of this theorem is also true and is similar to S. N. Bernstein's theorem for the periodic case (see § 3.8).

THEOREM 3 (V. K. Dzyadyk). *Suppose that we are given $r > 0$, and that the function $f(x)$ is defined on the segment $[a, b]$ and that for any $n = 0, 1, 2, \ldots$ there exists a polynomial $P_n(x) \in \mathfrak{P}_n$ such that for all $x \in [a, b]$*

$$|f(x) - P_n(x)| \leqslant \frac{c}{n^r}\left\{[\sqrt{(x - a)(b - x)}]^r + \frac{1}{n^r}\right\},$$

*where the constant $c$ does not depend on $n$ or $x$.*

*Then*

$$f(x) \in H^{(r)}.$$

Theorems 2 and 3 together give the necessary and sufficient condition for the function $f(x)$ to belong to the class $H^{(r)}$ on the segment $[a, b]$, which is expressed by the evaluation of the speed of approximation of the function $f(x)$ by polynomials of degree $n$, but as distinct from the periodic case these evaluations depend not only on the degree of the approximating polynomial, but also on the position of the point on the segment; by the same token in this condition there is no idea of the best approximation $E_{\mathfrak{P}_n}(f)$, although in this idea both conditions are rather closely connected.

The extensions of Theorems 1–3 with the help of the values of the deviations of the functions from their approximating polynomials has been obtained by A. F. Timan [19] using moduli of continuity.

† Let $r = \bar{r} + \alpha$, $\bar{r}$ is a non-negative whole number, $0 < \alpha \leqslant 1$. Theorem 2 with $0 < \alpha < 1$ has been proved by A. F. Timan, and with $\alpha = 1$ by V. K. Dzyadyk.

Note the inequalities for algebraic polynomials obtained by A. A. Markov and S. N. Bernstein which establish a connection between the absolute value of a polynomial and its derivative: let $P_n(x) \in \mathfrak{P}_n$ and

$$|P(x)| \leqslant L, \quad a \leqslant x \leqslant b,$$

then

$$|P'(x)| \leqslant \frac{2n^2 L}{b-a}, \quad a \leqslant x \leqslant b,$$

and

$$|P'(x)| \leqslant \frac{nL}{\sqrt{(b-x)(x-a)}}, \quad a < x < b.$$

These inequalities are exact in the sense that for any $n$ there exist polynomials for which they are transformed into an equality for at least one point.

Note, however, that for the proof of Theorem 3 a more exact inequality was required which was obtained by V. K. Dzyadyk: let the polynomial $P_n(x)$ of order not higher than $n$ satisfy, on the interval $(-1, 1)$ with an arbitrary real $\varrho$, the inequality

$$|P_n(x)| \leqslant L\left(\sqrt{1-x^2} + \frac{1}{n}\right)^{\varrho},$$

where $L$ is constant. Then for each $x \in (-1, 1)$ the derivative $P_n'(x)$ satisfies the inequality

$$|P_n'(x)| \leqslant cnL\left(\sqrt{1-x^2} + \frac{1}{n}\right)^{\varrho-1};$$

here $c$ is a constant which depends only on $\varrho$.

The property of being an infinitely, continuously differentiable function may be described in terms of best approximations as follows:

THEOREM 4 (D. Jackson). *Let $f(x) \in C^k[a, b]$ for all $k = 0, 1, \ldots$; then for any $k = 0, 1, \ldots$*

$$\lim_{n\to\infty} n^k E_{\mathfrak{P}_n}(f) = 0. \tag{3.15}$$

The theorem which in a sense is the converse of this reads as follows:

Theorem 5 (S. N. Bernstein). *Let the function $f(x)$ be defined on the segment $[a, b]$ and for any $k = 0, 1, 2, \ldots$ let*

$$\lim_{n\to\infty} n^k E_{\mathfrak{P}_n}(f) = 0;$$

*then the function $f(x) \in C[a, b]$ is infinitely, continuously differentiable on the interval $(a, b)$.*

In Theorem 5 only the differentiability on the interval $(a, b)$ was stated and therefore it is not a complete converse of Theorem 4.

We now turn to a characteristic of analytic functions using best approximations. A function $f(x)$, defined on the segment $[a, b]$, is said to be *analytic* on this segment if for each point $x_0$ of this segment there exists a neighbourhood $(x_0 - \delta, x_0 + \delta)$ at which the function $f(x)$ is given in the form of a power series, i.e.

$$f(x) = \sum_{n=0}^{\infty} a_n(x_0)\,(x - x_0), \quad x \in (x_0 - \delta, x_0 + \delta) \cap [a, b].$$

The aggregate of all functions which are analytic on the segment $[a, b]$ will be denoted by $A[a, b]$.

THEOREM 6 (S. N. Bernstein). *In order that $f(x)$ may belong to $A[a, b]$ it is necessary and sufficient that*

$$E_{\mathfrak{P}_n}(f) \leqslant kq^n,$$

*where $k$ and $q < 1$ are constants.*

*The condition*

$$\lim_{n \to \infty} \sqrt[n]{E_{\mathfrak{P}_n}(f)} = 0$$

*is necessary and sufficient for the function $f(x)$ to be an integral function.*

In terms of the theory of functions of a complex variable the properties of best approximations of analytic functions can be expressed as follows:

THEOREM 7 (S. N. Bernstein). *Let $f(x) \in C[a, b]$ and*

$$\overline{\lim}\ \sqrt[n]{E_{\mathfrak{P}_n}(f)} = \frac{1}{R} \quad (R > 1); \tag{3.16}$$

*then there exists a function $F(z)$ which is analytic inside the ellipse with foci at the points $a$ and $b$ with the sum of its semi-axes $\frac{b-a}{2} R$, which has a singular point on its contour and coincides with $f(x)$ on the segment $[a, b]$.*

*Conversely, if the function $F(z)$ is analytic inside the ellipse with foci at the points $a$ and $b$ and has a singular point on its contour, then its best approximation $E_{\mathfrak{P}_n}(F)$ on the segment $[a, b]$ satisfies the*

*condition* (3.16) *where* $\frac{b-a}{2}R$ *is the sum of the semi-axes of the given ellipse.*

For some classes of non-periodic functions $f(x)$ and also for some individual functions it is possible to establish not only the order of decrease of the best approximations $E_{\mathfrak{P}_n}(f)$ but also their exact or asymptotic value.

Let us consider first this problem for the individual function $f(x)$ and give an example of one of the theorems of S. M. Nikol'skii showing how the asymptotic value of $E_{\mathfrak{P}_n}(f)$ is determined by the individual properties of the function $f(x)$. Incidentally up to now we have not considered a problem of such a type.

S. N. Bernstein has shown that for any $k > 0$ there exists a finite limit

$$\lim_{n\to\infty} n^k E_{\mathfrak{P}_n}(|x|^k) = \mu(k),$$

i.e.

$$E_{\mathfrak{P}_n}(|x|^k) \approx \frac{\mu(k)}{n^k}, \quad n \to \infty.$$

THEOREM 8 (S. M. Nikol'skii). *Let $k$ be an odd number, $f(x) \in C^{(k-1}[-1, 1]$, the derivative $f^{(k-1)}(x)$ be absolutely continuous on $[-1, 1]$,† the derivative $f^{(k)}(x)$ have only discontinuities of the first kind on the segment $[-1, 1]$, and let there exist at least one $x_0 \in (-1, 1)$ such that $f^{(k)}(x_0 + 0) \neq f^{(k)}(x_0 - 0)$. Then*

$$E_{\mathfrak{P}_n}(f) \approx \frac{\varkappa\mu(k)}{2k!\,n^k} \quad \text{as} \quad n \to \infty,$$

*where*

$$\varkappa = \max_{-1<x<1} |f^{(k)}(x + 0) - f^{(k)}(x - 0)|(1 - x^2)^{\frac{k}{2}}.$$

A similar theorem was obtained by S. M. Nikol'skii for $k$ even; we need only substitute for $\mu(k)$ a constant

$$\nu(k) = \lim_{n\to\infty} n^k E_{\mathfrak{P}_n}(x|x|^{k-1}).$$

We turn now to asymptotic evaluations of best approximations for classes of functions.

† Then almost everywhere on $[-1, 1]$ there exists a derivative $f^{(k)}(x)$.

We shall represent by $\mathrm{Lip}_M 1$ the class of functions $f(x)$ which are defined on some fixed $[a, b]$ and which satisfy on it a Lipschitz condition with constant $M$:

$$|f(x+h) - f(x)| \leqslant M|h|, \quad x \in [a, b], \quad x + h \in [a, b].$$

THEOREM 9 (S. M. Nikol'skii). *The following is true*

$$\sup_{f \in \mathrm{Lip}_M 1} E_{\mathfrak{P}_n}(f) = \frac{M\pi}{2n} - \varepsilon_n,$$

*where* $\varepsilon_n > 0$, $\varepsilon_n = O\left(\frac{1}{n \ln n}\right)$.

Also let $\mathfrak{A}_K [a, b]$ be the class of functions which are analytic in the ellipse with foci at the points $a$ and $b$ and the sum of its semi-axes $h$, and in the given ellipse let

$$|\mathrm{Re}\, f(z)| \leqslant K.$$

We shall represent by $E_{\mathfrak{P}_n}(f)$ in this case the best approximation of the function $f(z) \in \mathfrak{A}_K [a, b]$ on the segment by means of polynomials of class $\mathfrak{P}_n$.

THEOREM 10 (N. I. Akhiezer).

$$\sup_{f(z) \in \mathfrak{A}_K [a, b]} E_{\mathfrak{P}_n}(f) = \frac{8K}{\pi} \sum_{l=0}^{\infty} (-1)^l \frac{\exp[-(2l+1)nh]}{1 + \exp[-2(2l+1)nh]} \frac{1}{2l+1}.$$

For other questions on the theory of uniform best approximations by algebraic polynomials see [1], [3], [4], [6], [7], [8], [9], [11] and [12].

**3.9. The best approximations of functions by integral functions.** Let us confine ourselves to the question of the approximation of functions which are defined and bounded on the whole axis. In this case algebraic polynomials are not suitable for the direct approximation of given functions because they are not bounded on the whole real axis. Here, therefore, it is necessary to take as approximating functions other integral functions of an exponential type.†

Let there be a fixed $\sigma > 0$. The function $g(z)$ of a complex argument $z$ is called *an integral function of exponential type* of degree not above $\sigma$ if

† For another approach to this question see § 3.10.

(1) $g(z)$ is expanded in a power series which converges on the whole plane,

(2) for any $\varepsilon > 0$ there exists a constant $A > 0$ such that for any $z$

$$|g(z)| < Ae^{(\sigma+\varepsilon)|z|}.$$

The aggregate of all integral functions of exponential type of degree not higher than $\sigma$, which are bounded on the whole real axis, we shall denote by $\mathfrak{B}_\sigma$.

The theory of best approximations of real functions, which are defined, bounded and continuous on the whole numerical axis, by means of integral functions of exponential type was built up by S. N. Bernstein and in a sense is similar to the theory of best approximations of periodic continuous functions by means of trigonometric polynomials. This analogy is expressed in particular by the fact that, as with trigonometric so for exponential functions, the following statement holds (S. N. Bernstein):

*If* $g(x) \in \mathfrak{B}_\sigma$, *then*

$$\sup_{-\infty<x<\infty} |g'(x)| \leqslant \sigma \sup_{-\infty<x<\infty} |g(x)|,$$

*where this inequality is exact, i.e. the sign of equality is attained, for functions of the form*

$$g(z) = ae^{i\sigma z} + be^{-i\sigma z}$$

*and for them only.*

In order to establish a connection between the degree of smoothness of the function considered and the speed of decrease of its best approximations by means of functions of class $\mathfrak{B}_\sigma$, it is useful to consider the class $H^{(r)}$ introduced in § 3.8 (in the definition of class $H^{(r)}$ in § 3.8 we must now consider $a = -\infty$, $b = \infty$); here in addition we shall assume that functions of class $H^{(r)}$ are bounded on the whole numerical axis.

The quantity

$$E_{\mathfrak{B}_\sigma}(f) = \inf_{g \in \mathfrak{B}_\sigma} \sup_{-\infty<x<\infty} |f(x) - g(x)|$$

is called *the best approximation of the function* $f(z)$ *by means of integral functions of class* $\mathfrak{B}_\sigma$.

THEOREM (S. N. Bernstein). *In order that the function* $f(x)$ *may belong to the class* $H^{(r)}$ *it is necessary and sufficient that there should exist a constant* $c > 0$ *such that for all integers* $n$

$$E_{\mathfrak{B}_\sigma}(f) \leqslant \frac{c}{n^r}. \tag{3.17}$$

Condition (3.17) remains sufficient for the function $f(z)$ to belong to class $H^{(r)}$ if it is satisfied not for all $n$ but only for a sequence of $n$ forming an increasing geometric sequence.

**3.10. Weighted approximations of functions on the whole numerical axis.** Let us now consider some questions of the approximation of continuous functions which are defined on the whole numerical axis, with respect to some given weight.

Let the function $\Phi(x)$ be defined and positive for all real $x$ and let

$$\lim_{|x|\to\infty} \frac{x^n}{\Phi(x)} = 0, \quad n = 1, 2, \ldots$$

We now consider the aggregate of all continuous functions, each of which is defined on the whole numerical axis and satisfies the condition

$$\lim_{|x|\to\infty} \frac{f(x)}{\Phi(x)} = 0.$$

This aggregate by means of the norm

$$\|f\|_\Phi = \sup_{-\infty<x<\infty} \frac{|f(x)|}{\Phi(x)} \tag{3.18}$$

is converted into normed space, which we may denote by $C^0_\Phi$. If the set of all algebraic polynomials is distributed densely everywhere in the space $C^0_\Phi$ with respect to the metric, induced by the norm (3.18), the function $\Phi(x)$ is said to be *weighted*. Thus the function $\Phi(x)$ is said to be weighted if any function of class $C^0_\Phi$ may be approximated by a polynomial with any degree of accuracy in the sense of the metric (3.18).

In order that the function $\Phi(x)$ may be weighted it is necessary and sufficient (S. N. Bernstein and N. I. Akhiezer) that

$$\sup_{P(x)\in\mathfrak{R}} \int_{-\infty}^{\infty} \frac{\ln|P(x)|}{1+x^2}\,dx < \infty,$$

where $\mathfrak{R}$ is the set of all algebraic polynomials $P(x)$ for which $\|P\|_\Phi \leqslant 1$.

S. N. Mergelyan showed that if

$$M(z) = \sup_{P\in L} |P(z)|,$$

where $L$ is the aggregate of all algebraic polynomials $P(x)$ for which

$$\sup_{-\infty<x<\infty} \frac{|P(x)|}{\Phi(x)\,(1+|x|)} \leqslant 1$$

the condition $M(z) = +\infty$ with $\operatorname{Im} z \neq 0$ is necessary and sufficient for the function $\Phi(x)$ to be weighted.

On the question of the approximation of functions with a weight [2], [3], [8] and [10] may be read.

## § 4. METHODS OF UNIFORM APPROXIMATION OF FUNCTIONS

In the present paragraph we shall consider some effective methods of forming polynomials which approximate a given function, and of evaluating the errors of the approximations of the functions, which are obtained by these methods. We have noted above various results connected with these questions (see Theorems 3 and 4 in §§ 3.3 and 3.6).

**4.1. The approximation of periodic functions by Fourier sums.** We shall represent the aggregate of all functions which are summable† on the segment $[-\pi, \pi]$ by $L[-\pi, \pi]$. Let $f(x) \in L[-\pi, \pi]$; then the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \tag{4.1}$$

for which the coefficients $a_k$ and $b_k$, called *the Fourier coefficients*, are determined by the formulae

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\,dx,$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx\,dx,$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx\,dx,$$

† For the definition of summability see Chapter I of this book.

as is well known, is called *the Fourier series* of the given function $f(x)$; its partial sum of the $n$th order is called *the Fourier sum of the nth order* of the function $f(x)$ and is denoted by $S_n(x; f)$.

If the function $f(x)$ has a period $2\pi$ its partial Fourier sums may be written in the form

$$S_n(x; f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(u) f(x + u)\, du,$$

where

$$D_n(u) = \frac{\sin \frac{2n + 1}{2} u}{\sin \frac{u}{2}}.$$

The function $D_n(u)$ is called the *Dirichlet kernel of the nth order.* It satisfies the following relations

$$D_n(-u) = D_n(u),$$

$$\frac{1}{\pi} \int_{0}^{\pi} D_n(u)\, du = 1.$$

THEOREM 1. *If* $f(x) \in C[-\pi, \pi]$, $g(x) \in C[-\pi, \pi]$ *and all the Fourier coefficients of the functions* $f(x)$ *and* $g(x)$ *coincide then the function* $f(x)$ *is identical with the function* $g(x)$.

THEOREM 2. *If the Fourier series of the function* $f(x) \in C^*_{2\pi}$ *converges at the point* $x_0$ *then it converges to* $f(x_0)$.

Note also Du Bois Reymond's result that for an arbitrary, continuous, periodic function $f(x)$ of period $2\pi$; its Fourier series is not bound to converge at all points of the period. A. N. Kolmogorov has constructed an example of a summable function whose Fourier series is divergent at all points.

THEOREM 3. *Let the function* $f(x)$ *be defined on the segment* $[-\pi, \pi]$ *and have bounded variation on it.*† *Then its Fourier series converges at each point* $x \in [-\pi, \pi]$ *to the value*

$$\frac{f(x - 0) + f(x + 0)}{2},$$

† Hence it follows that the function $f(x)$ is given as the difference between two bounded, monotonic functions and is thus integrable.

*and at the points* $x = \pm \pi$ *to the value*

$$\frac{f(-\pi + 0) + f(\pi - 0)}{2}.$$

In particular at points of continuity of the function $f(x)$ its Fourier series converges to $f(x)$.

We turn now to the question of the speed of approximation of the Fourier sums of a function $f(x)$ to the function itself.

THEOREM 4 (H. Lebesgue). *Let* $f(x) \in C^*_{2\pi}$, *then*

$$|S_n(x; f) - f(x)| \leqslant (3 + \ln n)\, E_{\mathcal{T}_n}(f).$$

This theorem shows that the smoother the function $f(x)$, in particular the more derivatives it has, the better its partial Fourier sums converge to it, because in this case, according to the results of § 3.7, the best approximations $E_{\mathcal{T}_n}(f)$ quickly decrease.

THEOREM 5. *Let the modulus of continuity* $\omega(\sigma; f)$ *of a periodic function* $f(x)$ *of period* $2\pi$ *satisfy the Dini–Lipschitz condition*

$$\lim_{\delta \to 0} \omega(\delta; f) \ln \delta = 0;$$

*then the function* $f(x)$ *can be expanded in a uniformly convergent Fourier series.*

The exact asymptotic expression for the deviation of the partial Fourier sums from functions of a given class was first obtained by A. N. Kolmogorov.

THEOREM 6 (A. N. KOLMOGOROV). *Let* $W^{(r)}_*(K)$ ($r$ *is a non-negative whole number*) *be the aggregate of all periodic functions* $f(x)$ *of period* $2\pi$ *which have a continuous derivative of order* $r$ *such that* $|f^{(r)}(x)| \leqslant K$. *Then*

$$\sup_{f(x) \in W^{(r)}_*(K)} \sup_{-\infty < x < \infty} |f(x) - S_n(x; f)| = \frac{4}{\pi^2} \frac{K \ln n}{n^r} + O\left(\frac{1}{n^r}\right).$$

If $r = 0$ this result is closely connected with one of H. Lebesgue's, viz. the inequality

$$M_n = \sup_{|f| < K} \sup_{-\infty < x < \infty} |S_n(x; f)| = \frac{4K}{\pi} \ln n + O(1),$$

$$f(x) \in C^*_{2\pi}.$$

We can show that from this or, to be more exact, from the fact that $\lim_{n \to \infty} M_n = \infty$, it follows that there exists a continuous

function whose Fourier series automatically diverges for at least one point of the period.

For the case of the function $f(x)$, whose $r$th derivative satisfies Hölder's condition, the corresponding result was obtained by S. M. Nikol'skii.

THEOREM 7 (S. M. Nikol'skii). *The following relation holds:*

$$\sup_{f(x)\in H^{(r)}_{*}(M)} \sup_{-\infty<x<\infty} |f(x) - S_n(x; f)| = \frac{2^{\alpha+1} M}{\pi \bar{r}} \frac{\ln n}{n^r} \int_0^{\pi/2} v^{\alpha} \sin v \, dv + O\left(\frac{1}{n^r}\right),$$

*where $\bar{r}$ is a non-negative whole number and $r = \bar{r} + \alpha$, $0 < \alpha < 1$* (see § 3.7).

Theorems 6 and 7 show that although the speed with which the Fourier sums approach the functions of a given class (i.e. the order of deviation of $S_n(x; f)$ from $f(x)$ for the given class as a whole) is in a sense less satisfactory than the speed of approach of the best approximations, for a series of problems (and above all those of an applied nature) this deficiency has no essential significance (the order of deviation of the Fourier sum differs from the order of the best approximations in the respective cases by not more than $\ln n$ as $n \to \infty$). This circumstance, and also the simplicity of the calculations of the Fourier sums and the fact that the speed of their convergence to the function is increased with increased smoothness of the function, makes for a wide application of the approximation of functions by means of Fourier sums both in theoretical and in applied questions. Fourier sums undoubtedly constitute one of the most convenient and widely used methods of function approximation.

**4.2. Linear methods of approximation of periodic functions by trigonometric polynomials (the methods of Féjer, de la Vallée-Poussin and Bernstein–Rogosinski).** Let $f(x) \in L[-\pi, \pi]$ and $S_n(x; f)$ be its Fourier sums of order $n = 0, 1, 2, \ldots$ The sum

$$\sigma_n(x; f) = \frac{S_0(x; f) + S_1(x; f) + \cdots + S_{n-1}(x; f)}{n}$$

is called *the Féjer sum* of the function $f(x)$ *of order* $n$.

In the case where $f(x)$ is a periodic function of period $2\pi$ the Féjer sum $\sigma_n(x; f)$ may be given in the form

$$\sigma_n(x; f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_n(u) f(x + u)\, du,$$

where

$$\Phi_n(u) = \frac{\sin^2 \frac{n}{2} u}{n \sin^2 \frac{u}{2}}.$$

The function $\Phi_n(u)$ is *the Féjer kernel* and possesses the following properties:

(1) $\Phi_n(u) \geqslant 0$,

(2) $\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_n(u)\, du = 1$,

(3) $M(\delta) \equiv \max_{0 < \delta \leqslant |u| \leqslant \pi} \Phi_n(u) \to 0 \quad \text{as} \quad n \to \infty$

with any fixed $\delta > 0$.

For functions $f(x) \in C^*_{2\pi}$ their Féjer sums satisfy the inequality

$$|\sigma_n(x; f)| \leqslant \max_{-\infty < x < \infty} |f(x)|.$$

THEOREM 1 (L. Féjer). *Let $f(x) \in C^*_{2\pi}$. Then the sequence of Féjer sums $\sigma_n(x; f)$ converges uniformly on the whole numerical axis to the function $f(x)$.*

The speed of approximation of the Féjer sums $\sigma_n(x; f)$ to the function $f(x)$ was originally investigated by S. N. Bernstein, then for a series of important classes of functions S. M. Nikol'skii found its exact asymptotic expression.

THEOREM 2 (S. M. Nikol'skii). *Let*

$$f(x) \in H^{(r)}_*(M), \quad 0 < r < 1,$$

*then*

$$\sup_{f \in H^{(r)}_*(M)} \max_{-\infty < x < \infty} |\sigma_n(x; f) - f(x)| = \frac{2M\Gamma(r)}{\pi(1 - r)} \frac{\sin \frac{r\pi}{2}}{n^r} + O\left(\frac{1}{n^r}\right).$$

THEOREM 3 (S. M. Nikol'skii). *Let*

$$f(x) \in \mathrm{Lip}_M^* 1,$$

*then*

$$\sup_{f \in \mathrm{Lip}_M^* 1} \max_{-\infty < x < \infty} |\sigma_n(x; f) - f(x)| = \frac{2M \ln n}{\pi n} + O\left(\frac{\ln n}{n}\right).$$

It is curious to note that in considering classes of smoother functions, for example classes $H_*^{(r)}$ with $r > 1$, the upper bound of the deviations of the Féjer sums of the $n$th order from the respective functions is not improved and cannot have an order higher than $1/n$. This is already obvious from the fact that for the integral function $f(x) = 1 - \cos x = 2 \sin^2 \frac{x}{2}$ we have

$$\sigma_n(0) = \frac{4}{n\pi} \int_0^{\pi/2} \sin^2 nt \, dt = \frac{1}{n}.$$

Sums of the form

$$V_{n,p} = \frac{S_n(x; f) + \cdots + S_{n+p-1}(x; f)}{p}$$

are called *the Vallée-Poussin sums of the* $(n, p)$*th order* for the function $f(x)$. We shall consider the simplest case where $p = n$ and write $V_{n,n}(x; f) = V_n(x; f)$. If the function $f(x)$ is a trigonometric polynomial of degree $m$, then for $n \geqslant m$ the Vallée-Poussin sum $V_n(x; f)$ coincides with $f(x)$ (these properties are obviously possessed by the Fourier sums but not by the Féjer sums). In addition

$$|V_n(x; f)| \leqslant 3 \max_{-\infty < x < \infty} |f(x)|, \quad f(x) \in C_{2\pi}^*.$$

THEOREM 4 (C. de la Vallée-Poussin). *Let* $f(x) \in C_{2\pi}^*$. *Then*

$$|V_n(x; f) - f(x)| \leqslant 4 E_{\mathfrak{T}_n}(f).$$

Let the function $f(x)$ be summable on the segment $[-\pi, \pi]$. The integral

$$J_n(x; f) = \frac{(2n)!!}{(2n-1)!!} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \cos^{2n} \frac{t - x}{2} \, dt$$

is called *the Vallée-Poussin singular integral of the nth order* for the function $f(x)$. The function $J_n(x; f)$ is a trigonometric polynomial of degree not higher than $n$.

THEOREM 5 (C. de la Vallée-Poussin). *Let* $f(x) \in C^*_{2\pi}$. *Then the sequence of its Vallée-Poussin integrals* $\{J_n(x; f)\}$ *converges uniformly on the whole numerical axis to the function* $f(x)$; *and here*

$$|J_n(x; f) - f(x)| \leqslant 3\omega\left(\frac{1}{\sqrt{n}}; f\right).$$

I. P. Natanson investigated the asymptotic behaviour of the deviations of the Vallée-Poussin integral on the whole from functions which satisfy Hölder's conditions. For convenience of formulation we assume according to definition that $\mathrm{Lip}^*_M \alpha = H^{(\alpha)}_*(M)$ if $0 < \alpha < 1$.

THEOREM 6 (I. P. Natanson).

$$\sup_{f(x) \in \mathrm{Lip}^*_M \alpha} \max_x |J_n(x; f) - f(x)| = \frac{2^\alpha M}{\sqrt{\pi n^\alpha}} \Gamma\left(\frac{1+\alpha}{2}\right) + O\left(\frac{1}{\sqrt{n^\alpha}}\right).$$

The Vallée-Poussin integral, although it approximates any continuous function, but on the whole with rather poor accuracy (in comparison with other methods, e.g. by means of best approximating polynomials), does so in the same way as do Bernstein polynomials or Féjer sums. We can show that no improvement of the smoothness of the function, for instance the increase of the exponent $r$ for classes $H^{(r)}_*$, leads for this class as a whole to a higher order of decrease of the above value than $1/n$.

As before let $S_n(x; f)$ be the Fourier sums of the function $f(x)$ and let us assume

$$B^*_n(x; f) = \frac{1}{2}[S_n(x + \alpha_n; f) + S_n(x - \alpha_n; f)],$$

where $\alpha_n > 0$ $(n = 1, 2, \ldots)$ and $\lim_{n\to\infty} \alpha_n = 0$.

The approximation of periodic functions by trigonometric polynomials $B^*_n(x; f)$, generally called *approximation by the Bernstein–Rogosinski method*, has been treated in the works of W. Rogosinski, S. N. Bernstein, A. F. Timan, S. B. Stechkin, F. I. Kharshiladze and others.

Note for example the following result [11]:

THEOREM 7. *Let* $f(x) \in C^*_{2\pi}$ *and* $\alpha_n = \dfrac{\pi}{2n+1}$. *Then*

$$|B^*_n(x; f) - f(x)| \leqslant (2\pi + 1) E_{\mathfrak{T}_n}(f) + \omega\left(\frac{2\pi}{2n+1}; f\right).$$

The linear methods of approximation of periodic functions by means of trigonometric polynomials and the greatest possible extensions of these methods (one of the most important of them is the method of approximation by singular integrals) indicate at the present time one of the most developed chapters in the theory of functions. Great interest in particular is attached to the study of the best linear methods of approximation for a given class of functions. Information on all these questions may be found in [1], [3], [7], [8], [11], [12], and [13].

**4.3. Linear methods of approximation of functions by algebraic polynomials.** If the function $f(x)$ has derivatives of all orders and is also analytic, the natural method of its approximation is by means of the partial sums of a Taylor series (see § 1.2). The study of the class of analytic functions is, however, simpler in a complex than in a real domain. Therefore we shall assume that the function $f(z)$ considered is defined in the unit circle $|z| \leqslant 1$ of the complex plane.

We represent by $A^{(r)}(K)$ $(r = 0, 1, \ldots)$ the aggregate of all functions $f(z)$ which are analytic in the circle $|z| < 1$ and satisfy the condition

$$|f^{(r)}(z)| \leqslant K, \quad |z| < 1.$$

Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$, $f(z) \in A^{(r)}$ $(k)$ with some fixed $r = 0, 1, 2, \ldots$ We assume

$$P_n(z; f) = \sum_{k=0}^{n} a_k z^k,$$

$$r_n(z; f) = f(z) - P_n(z; f) = \sum_{k=n+1}^{\infty} a_k z^k,$$

$$R_n(f) = \max_{|z|<1} |r_n(z; f)|,$$

$$R_n[A^{(r)}(K)] = \sup_{f \in A^{(r)}(K)} R_n(f).$$

The following proposition is true:

THEOREM 1 (S. B. Stechkin). *Let $r$ be a natural number and $f(z) \in A^{(r)}(K)$. Then*

$$r_n(z; f) = -\frac{Kz^r}{(n+1)^r} P_{n-r}(z; f^{(r)}) + O\left(\frac{1}{(n+1)^r}\right), \quad r \leqslant n.$$

(This formula is realized uniformly in the circle $|z| < 1$ and uniformly with respect to every class $A^{(r)}(K)$.)

THEOREM 2 (S. B. Stechkin).

$$R_n[A^{(r)}(K)] = \frac{K}{\pi}\frac{\ln(1+n)}{(1+n)^r} + O\left(\frac{1}{(n+1)^r}\right), \quad n \geqslant r-1.\dagger$$

The above value is attained in class $A^{(r)}(K)$ in the sense that there exists a function $f_0(z) \in A^{(r)}(K)$ such that

$$\overline{\lim_{n\to\infty}} \frac{|r_n(1; f_0)|}{\frac{K}{\pi}\frac{\ln(1+n)}{(1+n)^r}} = 1.$$

For non-analytic functions on an analogy with periodic we consider their expansion in a generalized Fourier series with respect to some system of orthogonal polynomials (with weight). In more detail: suppose we are given a system $\{p_n(x)\}$ of polynomials, $n = 0, 1, \ldots$, which are orthogonal with weight $q(x) \geqslant 0$ on the segment $[a, b]$ i.e.

$$\int_a^b q(x)\, p_n(x)\, p_m(x)\, dx \begin{cases} > 0 & \text{if } n = m, \\ = 0 & \text{if } n \neq m, \end{cases}$$

and also suppose that on the segment $[a, b]$ we are given a (summable) function $f(x)$. Then the series

$$\sum_{n=0}^{\infty} a_n p_n(x), \tag{4.2}$$

whose coefficients $a_n$, the so-called *Fourier coefficients* of the function $f(x)$ *with respect to the system* $\{p_n(x)\}$, are determined by the formula

$$a_n = \frac{\int_a^b q(x)\, f(x)\, p_n(x)\, dx}{\int_a^b q(x)\, p_n^2(x)\, dx}$$

† In the case where $r = 0$ this value is closely connected with the following inequality of E. Landau:

$$\sup_{\varphi \in B^{(0)}} \max_{|z| \leqslant 1} |P_n(z, \varphi)| = \frac{1}{\pi} \ln(1+n) + O(1), \quad n \geqslant 0.$$

is called the Fourier series of the function $f(x)$ with respect to the system $\{p_n(x)\}$.

As an example of such expansions (and of approximations of functions by the partial sums of series (4.2)) we shall consider an expansion in Legendre and Chebyshev polynomials.

The polynomials $P_n(x)$, defined by the formulae

$$P_0(x) = 1, \quad P_n(x) = \frac{1}{2^n n!} \frac{d^n(x^2 - 1)^n}{dx^n},$$

are called *Legendre polynomials*. On the segment $[-1, 1]$ they form an orthogonal system

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx = \begin{cases} 0, & \text{if } m \neq n, \\ \dfrac{2}{2n+1} & \text{if } m = n. \end{cases}$$

The Fourier coefficients $a_n$, $n = 0, 1, \ldots$ of the function $f(x)$ with respect to the system of Legendre polynomials are determined by the formula

$$a_n = \frac{2n+1}{2} \int_{-1}^{1} f(x)\, P_n(x)\,dx.$$

The convergence of a series of Legendre polynomials is characterized in the following way:

THEOREM 3. *Suppose that on the segment* $[-1, 1]$ *the function* $f(x)$ *satisfies the Dini–Lipschitz condition*

$$\lim_{\delta \to 0} \omega(\delta; f) \ln \delta = 0;$$

*then at all points of the interval* $(-1, 1)$ *the Fourier series of the function* $f(x)$ *with respect to the system of Legendre polynomials converges to the function* $f(x)$, *and this uniformly, on any segment* $[-1 + h, 1 - h]$, $h > 0$.

THEOREM 4. *Let the function* $f(x)$ *be defined on the segment* $[-1, 1]$ *and let*

$$\lim n^2 E_{\mathfrak{P}_n}(f) = 0; \tag{4.3}$$

*then the Fourier series of the function* $f(x)$ *with respect to the system of Legendre polynomials converges uniformly to the function* $f(x)$ *itself on the segment* $[-1, 1]$.

Note that, according to the results given in § 3.9, the condition (4.3) is automatically satisfied if the function $f(x)$ is twice continuously differentiable on the segment $[-1, 1]$.

Now suppose we are given a system of Chebyshev polynomials

$$T_0(x) = 1, \; T_n(x) = \frac{3}{2^{n-1}} \cos (n \arccos x);$$

these form an orthogonal system with weight $\frac{1}{\sqrt{1-x^2}}$ on the segment $[0, 1]$:

$$\int_{-1}^{1} \frac{T_n(x)\, T_m(x)}{\sqrt{1-x^2}}\, dx = \begin{cases} 0, & \text{if } m \neq n, \\ \dfrac{\pi}{2^n}, & \text{if } m = n. \end{cases}$$

The partial sums of the Fourier series of the function $f(x)$ with respect to the Chebyshev polynomials

$$\sum_{n=0}^{n} a_n T_n(x)$$

will be denoted by $U_n(x; f)$

$$U_n(x; f) = \sum_{k=0}^{n} a_n T_n(x),$$

$$a_n = \frac{2^n}{\pi} \int_{-1}^{1} \frac{f(x)\, T_n(x)}{\sqrt{1-x^2}}\, dx.$$

THEOREM 5. *Let the function $f(x)$ be defined on the segment $[-1, 1]$ and let*

$$\lim_{n \to \infty} E_{\mathfrak{P}_n}(f) \ln n = 0; \tag{4.4}$$

*then the Fourier series of the function $f(x)$ with respect to the system of Chebyshev polynomials converges uniformly to the function $f(x)$ itself on the segment $[-1, 1]$, and also*

$$|U_n(x; f) - f(x)| \leqslant (3 + \ln n)\, E_{\mathfrak{P}_n}(f).$$

Note also that condition (4.4) is automatically fulfilled (according to the results of § 3.9) if the function $f(x)$ satisfies a Dini–Lipschitz condition on the segment $[-1, 1]$. It can be shown that the results cited here are in the defined sense exact, whence it follows that, for uniform convergence of the Fourier series with

respect to the Chebyshev polynomials, there is rather less smoothness than for the uniform convergence of the Fourier series with respect to Legendre polynomials.

As in the periodic case, with the approximation of functions by partial sums of Fourier series with respect to some orthogonal systems of polynomials, in a series of cases we can also evaluate the upper bounds of the deviations of the partial sums of these series from the functions themselves.

Let $r$ be a fixed non-negative whole number.

We denote by $W^{(r)}(K)$ the aggregate of functions $f(x)$ which are defined and continuous on the segment $[-1, 1]$ together with their derivatives up to order $r$ inclusive and such that

$$|f^{(r)}(x)| \leqslant K, \quad -1 \leqslant x \leqslant 1.$$

For the partial sums of the expansions of the functions in Chebyshev polynomials we have:

THEOREM 6 (S. G. Selivanova).

$$\sup_{f(x) \in W^{(r)}(K)} |f(x) - U_n(x; f)| = \frac{4K}{\pi^2} \frac{\ln n}{n^r} (1 - x^2)^{\frac{r}{2}} + O\left(\frac{1}{n^r}\right).$$

Again we see that here, as distinct from the periodic case, the right-hand side depends on the position of the point $x$ on the segment $[-1, 1]$.

An account of other questions connected with linear methods of approximation of functions by algebraic polynomials may be found in [1], [3], [7], [8], [11] and [12].

**4.4. The best linear methods of approximation of functions.** Suppose that we are given a Banach space $R$ of functions $f(x)$ whose norm as usual is denoted by $\|f\|$, and further that we are given some method of approximation of the function $f(x) \in R$, or to put it in another way are given a sequence of operators $\{U_n\}_{n=1}^{\infty}$

$$U_n(f) \in R, \quad f(x) \in R, \quad n = 1, 2, \ldots$$

The method of approximation $\{U_n\}_{n=1}^{\infty}$ is said to be linear if for any two functions $f_1(x) \in R$ and $f_2(x) \in R$ and any numbers $\lambda_1$ and $\lambda_2$

$$U_n(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 U_n(f_1) + \lambda_2 U_n(f_2).$$

Now let there be given in the space $R$ a system of functions

$$\Phi = \{\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n, \ldots\}.$$

Let us consider the problem of the approximation (in the sense of the norm of space $R$) of a given fixed function $f(x) \in R$ with the help of all possible linear aggregates of the form $c_0\varphi_0 + c_1\varphi_1 + \dots + c_n\varphi_n$ ($n$ is fixed, $n = 0, 1, 2, \dots$) and let us assume that

$$E_{\Phi_n}(f) = \inf_{\substack{-\infty < c_i < \infty \\ i=0,1,\dots,n}} \|f - \sum_{i=0}^{n-1} c_i\varphi_i\|.$$

Now let $R^*$ be some subset of the space $R$ and let us assume

$$E_{\Phi_n}(R^*) = \sup_{f \in R^*} E_{\Phi_n}(f).$$

We say that the linear method of approximation $\{U_n\}_{n=1}^{\infty}$ is the *best* linear method of approximation of the functions of the set $R^*$ *with respect to the system* $\Phi$ if

$$E_{\Phi_n}(R^*) = \sup_{f \in R^*} \|f - U_n(f)\|, \quad n = 0, 1, 2, \dots$$

We also say that the linear method $\{U_n\}_{n=0}^{\infty}$ is the *asymptotically best* linear method of approximation of functions of the set $R^*$ *with respect to the system* $\Phi$ if

$$\lim_{n\to\infty} \frac{E_{\Phi_n}(R^*)}{\sup_{f \in R^*} \|f - U_n(f)\|} = 1.$$

The best linear methods have been investigated for the periodic case in the works of J. Favard, N. I. Akhiezer, M. G. Krein, B. Nagy, S. M. Nikol'skii, V. K. Dzyadyk, S. B. Stechkin and others.

THEOREM 1 (J. Favard, N. I. Akhiezer and M. G. Krein). *Let* (see § 3.7)

$$f(x) \in W_*^{(\Psi)}(K),$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx,$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx, \quad n = 1, 2, \dots$$

*be the coefficients of the Fourier series of a function $f(x)$. Then there exist numbers $\lambda_k^{(n)}$, $k = 1, 2, \ldots, n-1$; $n = 1, 2, \ldots$ such that the method of approximation*

$$U_n(x; f) = \frac{a_0}{2} + \sum_{k=1}^{n-1} \lambda_k^{(n)}(a \cos kx + b_k \sin kx)$$

*is the best (obviously linear) method of approximation of functions of the class $W_*^{(r)}(K)$ with respect to the system $\mathfrak{T}_{n-1}$ of trigonometric polynomials of degree not higher than $n-1$.*

Here the values of the coefficients of summability $\lambda_k^{(n)}$ can be effectively calculated. They have a particularly simple form if $r = 1$:

$$\lambda_k^{(n)} = \frac{k\pi}{2n} \cot \frac{k\pi}{2n}, \quad k = 1, 2, \ldots, \quad n-1, \quad n = 1, 2, \ldots$$

As we have remarked earlier (see Theorem 10, § 3.7) for the class considered, $W_*^{(r)}(K)$, J. Favard, and then N. I. Akhiezer and M. G. Krein, established an accurate value for the quantity $E_{\mathfrak{T}_n}(W_*^{(r)}(K))$ and then, in view of Theorem 1, also an accurate value for the quantity

$$\sup_{f(x) \in W_*^{(r)}(K)} \max_x |f(x) - U_n(x; f)|.$$

For the non-periodic case note S. M. Nikol'skii's theorem on one best asymptotic, linear method of approximation of functions by algebraic polynomials.

THEOREM 2 (S. M. Nikol'skii). *Let $f(x) \in \mathrm{Lip}_M 1$, $x \in [-1, 1]$ (see § 3.8), $P_0(x) = 1$, $P_k(x) = \cos k \arccos x$ are Chebyshev polynomials ($k = 1, 2, \ldots$)*

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{f(x)}{\sqrt{1 - x^2}}\, dx,$$

$$a_k = \frac{2}{\pi} \int_{-1}^{1} \frac{f(t)\, P_k(t)}{\sqrt{1 - t^2}}\, dt, \quad k = 1, 2, \ldots,$$

$$\lambda_k^{(n)} = \frac{k\pi}{2n} \cot \frac{k\pi}{2n}, \quad k = 1, 2, \ldots, \quad n-1, \quad n = 1, 2, \ldots,$$

*and let*

$$U_n(x; f) = \frac{a_0}{2} + \sum_{k=1}^{n-1} \lambda_k^{(n)} a_k P_k(x). \tag{4.5}$$

*Then*

$$\frac{M\pi}{2n} - \varepsilon < \sup_{f(x) \in \mathrm{Lip}_M 1} \max_{-1 \leqslant x \leqslant 1} |f(x) - U_n(f; x)| < \frac{M\pi}{2n},$$

*where*

$$\varepsilon_n = O\left(\frac{1}{n \ln n}\right) \text{ and } \varepsilon_n > 0.$$

Then the method of approximation (4.5) is the asymptotically best linear method of approximation of the functions $f(x)$ of class $\mathrm{Lip}_M 1$, $x \in [-1, 1]$ by polynomials of degree not higher than $n - 1$.

## § 5. THE APPROXIMATION OF FUNCTIONS OF ONE VARIABLE IN THE MEAN

**5.1. General remarks.** We shall now consider questions of approximation in the mean for functions which are summable in some degree. Let $E$ be a measurable set on a straight line $p \geqslant 1$, $q(x)$ a summable, almost everywhere positive function on $E$ (generally called a weight). We shall consider functions $f(x)$ which are defined and measurable on the set $E$. The aggregate of all measurable functions $f(x)$, for which the integral $\int q(x)\,|f(x)|^p\,dE$ is finite, will be denoted by $L_{p,q(x)}(E)$ and the number

$$\|f\|_{L_{p,q(x)}(E)} = \left[\int q(x)\,|f(x)|^p\,dE\right]^{\frac{1}{p}}, \tag{5.1}$$

is called the *norm* of the function $f(x) \in L_{p,q(x)}(E)$. The aggregate of the functions $L_{p,q(x)}(E)$ for ordinary operations of addition of functions and their multiplication by a number, in virtue of the definition of the norm (5.1), form a linear normed space,† which we shall represent by the same symbol $L_{p,q(x)}(E)$ or for short $L_{p,q(x)}$. In the case where $q(x) \equiv 1$ on $E$ we shall represent the given space simply by the symbol $L_p(E)$ or $L_p$ respectively and the norm of the element $f(x) \in L_p$ by the symbol $\|f\|_p$.

† Here equivalent functions are considered to be identical elements of this space.

If for $p = 2$ in space $L_{2,q(x)}(E)$ we introduce the scalar product by the formula

$$(f, g) = \int q(x) f(x) g(x)\, dE,$$

$$f(x) \in L_{2,q(x)}(E), \quad g(x) \in L_{2,q(x)}(E),$$

the space $L_{2,q(x)}(E)$ is converted into Hilbert space. Let the system $\mathfrak{F} = \{f_1, f_2, \ldots, f_n, \ldots\}$ of elements $f_n \in \mathfrak{F}$ be an orthogonal system of the space $L_{2,q(x)}(E)$, i.e. $(f_i, f_j) = 0$ if $i \neq j$, $(f_i, f_i) > 0$, and let $y \in L_{2,q(x)}(E)$. Then the following is true:

THEOREM. *The quantity*

$$\left\| y - \sum_{k=1}^{n} a_k f_k \right\|_{L_{2,q(x)}(E)}$$

*attains its least value with*

$$a_k = \frac{(y, f_k)}{(f_k, f_k)}, \quad k = 1, 2, \ldots$$

The coefficients $a_k$, calculated by this formula, are called the *Fourier coefficients of the element y with respect to the given system* $\mathfrak{F}$, and the series $\sum_{k=1}^{\infty} a_k f_k$ *the Fourier series of the element y with respect to the system* $\mathfrak{F}$.

We have already met with Fourier series with respect to a trigonometrical system (see § 4.1) and with respect to Legendre and Chebyshev polynomials (see § 4.3). As one more example we note the Jacobian polynomial

$$J_0^{(\alpha,\beta)} = 1,$$

$$J_n^{(\alpha,\beta)}(x) = K_n(1 - x)^{-\alpha}(1 + x)^{-\beta} \frac{d^n}{dx^n}[(1 - x)^{\alpha+n}(1 + x)^{\beta+n}],$$

where $K_n$ $(n = 1, 2, \ldots)$ are constants, $\alpha > -1, \beta > -1$. Jacobian polynomials form an orthogonal system with weight $p(x) = (1 - x)^{\alpha}(1 + x)^{\beta}$ on the segment $[-1, 1]$; if $\alpha = \beta = 0$ the Jacobian polynomials become Legendre polynomials, and if $\alpha = \beta = -\frac{1}{2}$ Chebyshev polynomials (with corresponding change of the constants $K_n$, $n = 1, 2, \ldots$).

For the case of the infinite interval $[0, \infty)$ the so-called *Laguerre polynomials*

$$L_n(x) = e^x \frac{d^n}{dx^n}(x^n e^{-x}), \quad n = 0, 1, \ldots,$$

form an orthogonal system with weight $q(x) = e^{-x}$ and on the whole numerical axis the *Hermite polynomials*

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n e^{-x}}{dx^n}, \quad n = 0, 1, 2, \ldots,$$

form an orthogonal system with weight $q(x) = e^{-x^2}$.

Note that in the case where mes $E < \infty$

$$\lim_{p \to \infty} \|f\|_p = \sup_{x \in E} \text{vrai} \, |f(x)|.$$

Therefore it is natural to assume by definition

$$\|f\|_\infty = \sup_{x \in E} \text{vrai} \, |f(x)|.$$

Linear normed space of functions with finite norm $\|f\|_\infty$ will be represented by $L_\infty(E)$ or simply $L_\infty$.

**5.2. Best approximations in the mean.** To spaces of functions of the type $L_p$ $(1 \leqslant p \leqslant \infty)$ we may carry over many of the results of the theory of best uniform approximations, in particular the results which establish a connection between the smoothness of functions and the order of the speed of decrease of their best approximations.

Let $E$ be an interval on a straight line, which is finite or infinite, and let the function $f(x)$ be defined on $E$. If the function $f(x)$ is absolutely continuous on $E$ then almost everywhere the function $f(x)$ has a derivative $f'(x)$. We shall call any function, which is defined on the set $E$ and almost everywhere coincides with $f'(x)$, *a generalized derivative of the function* $f(x)$ and represent it by the same symbol $f'(x)$. *A generalized derivative* $f^{(k)}(x)$ *of order* $k$ of the function $f(x)$ is a generalized derivative of the ordinary derivative $f^{(k-1)}(x)$ of order $k - 1$ of the function $f(x)$. If $f'(x)$ is a generalized derivative of the function $f(x)$, then

$$f(x) = f(a) + \int_a^x f'(x)\, dt, \quad a \in E.$$

Now let $1 \leqslant p \leqslant \infty$, $r = \bar{r} + \alpha$, $\bar{r}$ is a non-negative integer, $0 < \alpha \leqslant 1$, $M > 0$, $E = (-\infty, \infty)$. The function $f(x)$ is called a function of class $H_p^{(r)}(M)$ if

(1) $f(x) \in L_p\ (-\infty, \infty)$

(2) there exists a generalized derivative

$$f^{(\bar{r})}(x) \in L_p\ (-\infty, \infty)$$

such that

$$\| f^{(\bar{r})}(x + h) - 2f(x) + f^{(\bar{r})}(x - h) \|_p \leqslant M |h|^{\alpha}.$$

Similarly if $E = [0, 2\pi]$, then retaining the rest of the assumptions we shall call a periodic function $f(x)$ with period $2\pi$ a function of class $H_{p\,*}^{(r)}(M)$ if

(1) $f(x) \in L_p\,[0, 2\pi]$,

(2) there exists a generalized derivative

$$f^{(\bar{r})}(x) \in L_p\,[0, 2\pi]$$

such that

$$\| f^{(\bar{r})}(x + h) - 2f^{(\bar{r})}(x) + f^{(\bar{r})}(x - h) \|_p < M |h|^{\alpha}. \qquad (5.2)$$

In the case where $0 < \alpha < 1$ the condition (5.2) in the definition of classes $H_p^{(r)}(M)$ and $H_{p\,*}^{(r)}(M)$ is equivalent to the simpler condition

$$\| f^{(\bar{r})}(x + h) - f^{(\bar{r})}(x) \|_p \leqslant K |h|^{\alpha},$$

where $K > 0$ is a constant. As usual we assume that

$$H_p^{(r)} = \bigcup_{M>0} H_p^{(r)}(M) \quad \text{and} \quad H_{p\,*}^{(r)} = \bigcup_{M>0} H_{p\,*}^{(r)}(M).$$

We shall represent by $E_{\mathfrak{T}_n}^p(f)$ the best approximation in space $L_p$ of the function $f(x)$ by trigonometric polynomials of class $\mathfrak{T}_n$ on the segment $[0, 2\pi]$ i.e.

$$E_{\mathfrak{T}_n}^p(f) = \sup_{T(x) \in \mathfrak{T}_n} \| f - T \|_p.$$

We shall, moreover, represent by $\mathfrak{G}_\nu^p (\nu > 0)$ the class of integral functions of exponential type, of degrees not exceeding $\nu$ and summable in the $p$th degree in the whole numerical axis. We can show that (see § 3.8)

$$\mathfrak{G}_\nu^p \subset \mathfrak{G}_\nu^\infty = \mathfrak{G}_\nu, \quad 1 \leqslant p < \infty.$$

The best approximation of the function $f$, defined on the whole axis by means of functions of class $\mathfrak{G}_\nu^p$ in the sense of space $L_p$, we shall denote by $E^p_{\mathfrak{G}_\nu}(f)$, i.e.

$$E^p_{\mathfrak{G}_\nu}(f) = \sup_{g(x) \in \mathfrak{G}_\nu^p} \|f - g\|_p.$$

THEOREM 1. *In order that* $f(x) \in H^{(r)}_{p*}$, $1 \leqslant p \leqslant \infty$, $r > 0$, *it is necessary and sufficient that there should exist a constant* $c > 0$ *such that*

$$E^p_{\mathfrak{T}_\nu}(f) \leqslant \frac{c}{n^r}. \tag{5.3}$$

*Here the condition* (5.3) *remains sufficient for the function* $f$ *to belong to class* $H^{(r)}_{p*}$, *even where it is only satisfied for a sequence of integers* $n$ *which form an increasing geometric progression.*

THEOREM 2. *In order that* $f(x) \in H^{(r)}_p$, $1 \leqslant p \leqslant \infty$, $r > 0$, *it is necessary and sufficient that there should exist a constant* $c > 0$ *such that*

$$E^p_{\mathfrak{G}_\nu}(f) \leqslant \frac{c}{\nu^r}. \tag{5.4}$$

*The condition* (5.4) *remains sufficient for the function* $f(x)$ *to belong to class* $H^{(r)}_p$ *even where it is satisfied only for a sequence of subscripts* $\nu$ *forming an increasing geometric progression.*

The question of the connection of best approximations, in the sense of space $L_p$ by algebraic polynomials with the smoothness of the function, is being studied much less fully at the present time.

Within the meaning of the metric of space $L_p$ various linear methods of approximation of functions have also been studied, in particular the best linear methods of approximation for different classes of functions, and also evaluations of the upper bounds of the deviations of the approximations from the functions of some classes have been obtained (in various cases these upper bounds of the deviations have even been calculated accurately or asymptotically). It must be noted that some questions produce their own simpler and even more complete solution in the metric $L_p$, frequently for $1 < p < \infty$, and the values $p = 1$ and $p = \infty$ are in this sense as it were singular.

Information about these results may be found in [1], [7], [8], [11], [12] and [13].

**5.3. Systems of functions which best approximate some class of functions.** In all the previous investigations, when we encountered questions of the approximation of periodic functions, we always used linear combinations of sines and cosines of multiple arcs for this purpose. The question naturally arises as to whether there are not other systems which essentially differ from that given, linear combinations of which would give generally speaking better approximations of periodic functions. In 1935 A. N. Kolmogorov posed a similar question in general form and put forward its solution for a number of important cases of both periodic and non-periodic functions. We shall limit ourselves here only to the formulation of A. N. Kolmogorov's result for the periodic case, as being the simpler. This result justifies from the defined point of view the approximation of periodic functions by linear combinations of sines and cosines of multiple arcs; the theorem formulated below shows that in this case the value of the approximation turns out to be least.

We now formulate A. N. Kolmogorov's general problem. Suppose we are given a normed space $R$ of elements $f$, whose norm we shall as usual denote by $\|f\|$. Let $\mathfrak{F}$ be some subset of the space $R$ and $n$ a fixed natural number. Let us assume that

$$D_n(\mathfrak{F}) = \inf_{\varphi_1, \ldots, \varphi_n} \sup_{f \in \mathfrak{F}} \inf_{c_1, \ldots, c_n} \left\| f - \sum_{i=1}^{n} c_i \varphi_i \right\|,$$
$$\varphi_i \in R, \quad -\infty < c_i < \infty, \quad i = 1, 2, \ldots, n.$$

The problem of the determination and attainment of $D_n(\mathfrak{F})$ is considered, i.e. the problem of the best approximation of the set $\mathfrak{F} \subset R$ by means of "polynomials" of $n$ terms, $c_1\varphi_1 + \cdots + c_n\varphi_n$, in space $R$.

In the case where the lower bound of $D_n(\mathfrak{F})$ is attained for some system of functions $\varphi_1, \ldots, \varphi_n$ there naturally arises the question of the uniqueness of the determination by this condition of such a system $\varphi_1, \ldots, \varphi_n$ (a unique determination is understood here of course with accuracy to within the linear transformation of the system). Obviously

$$D_1(\mathfrak{F}) \geqslant D_2(\mathfrak{F}) \geqslant \cdots \geqslant D_n(\mathfrak{F}) \geqslant \cdots$$

Now let $R = L_2^*$ and $\mathfrak{F} = W_2^{(r)}$ be the aggregate of periodic functions with period $2\pi$, which are $r$ times differentiable functions

($r = 0, 1, \ldots$) such that

$$\|f^{(r)}\|_2 = \left(\int_{-\pi}^{\pi} |f^{(r)}(x)|^2 dx\right)^{\frac{1}{2}} \leqslant 1.$$

THEOREM (A. N. Kolmogorov).

$$D_{2m-1}\left(W_2^{(r)}\right) = D_{2m}\left(W_2^{(r)}\right) = \left(\frac{1}{2\pi m}\right)^r,$$

*in the case where* $n = 2m + 1$, *the lower bound of* $D_n(W_2^{(r)})$ *being reached on the system of functions* $1, \cos x, \sin x, \ldots, \cos mx, \sin mx$ (*this system is determined uniquely by the conditions of the theorem to within the accuracy of the linear transformation*).

## § 6. THE APPROXIMATION OF FUNCTIONS OF MANY VARIABLES

**6.1. Basic ideas.** For the study of questions of the approximation of functions of many variables by polynomials and their generalizations as a rule there arise fundamentally new difficulties which have no analogy in the one-dimensional case. Because of this the multi-dimensional case has been studied to a much lesser extent than the one-dimensional. The investigation of approximations of functions of many variables by various methods is most interesting, not only in connection with those applications of the theory of approximation which are met also for functions of one variable, but also in connection with a number of new applications which have a multi-dimensional character. The theory of approximations of functions of many variables finds for instance very effective use in the study of the relation between properties of a function and the properties of its boundary values, and also in the study of methods of best, or in a sense most appropriate, extensions of functions (theorems of such a type are particular cases of the direct and converse so-called theorems of enclosure). These questions in their turn play a fundamental part in the theory of equations of mathematical physics.

One of these questions, which has been studied rather fully in the many-dimensional case, is the question of the relation between the order of decrease of the best approximations by trigonometric polynomials, or integral functions of the exponential type, and the smoothness of the function in some sense or other. Here we cite some direct and converse theorems in the form in which they

were obtained by S. M. Nikol'skii, and also some inequalities obtained by S. M. Nikol'skii which have a fundamental significance both for the theory of approximation itself and for its application (for example to the theory of enclosure). We now turn to the discussion of these results.

We shall call functions of the form

$$T_{m_1 \dots m_n}(x_1, \dots, x_n) = \sum_{k_1=-m_1}^{m_1} \cdots \sum_{k_n=-m_n}^{m_n} c_{k_1 \dots k_n} \exp i \sum_{s=1}^{n} k_s x_s$$

$$c_{k_1 \dots -k_j \dots k_n} = \bar{c}_{k_1 \dots k_j \dots k_n}$$

*real trigonometric polynomials* of degree not higher than $m_1, \dots, m_n$ respectively in the variables $x_1, x_2, \dots, x_n$, (see § 2.2). The class of all trigonometric polynomials $T_{m_1, \dots, m_n}$ will be represented by $\mathfrak{T}_{m_1, \dots, m_n}$.

The function $g_{\nu_1 \dots \nu_n}(z_1, \dots, z_n)$ is called *an integral function of exponential type* of degrees $\nu_1, \dots, \nu_n$ respectively in the variables $z_1, \dots, z_n$, if

(1) the function $g_{\nu_1 \dots \nu_n}(z_1, \dots, z_n)$ is expanded in a power series which converges for any complex $z_1, \dots, z_n$;

(2) for any $\varepsilon > 0$ there exists $A > 0$ such that for all $z_1, \dots, z_n$

$$|g_{\nu_1 \dots \nu_n}(z_1, \dots, z_n)| \leqslant A \exp \sum_{k=1}^{n} (\nu_k + \varepsilon) |z_k|.$$

The aggregate of all integral functions of exponential type of the form $g_{\nu_1 \dots \nu_n}(z_1, \dots, z_n)$, which are summable in the $p$th degree in the whole real $n$-dimensional space will be represented by $\mathfrak{G}_{\nu_1 \dots \nu_n}$.

We shall represent by $L_p^{*(n)}$ the space of functions which are measurable in $n$-dimensional space, periodic of period $2\pi$ with respect to each variable and summable in the $p$th degree over a period (i.e. for $0 \leqslant x_i \leqslant 2\pi$, $i = 1, 2, \dots, n$). We shall assume that for $f(x) \in L_p^{*(n)}$

$$\|f\|_p^{(n)} = \left( \int_0^{2\pi} \int_0^{2\pi} \cdots \int_0^{2\pi} |f(x_1, \dots, x_n)|^p \, dx_1 \dots dx_n \right)^{\frac{1}{p}}, \tag{6.1}$$

and in the case where $p = \infty$

$$\|f\|_\infty^{(n)} = \sup_{0 < x_i < 2\pi} \operatorname{vrai}_{i=1, 2, \dots, n} |f(x_1, \dots, x_n)|. \tag{6.2}$$

We shall denote by $L_p^{(n)}$ the space of measurable functions which are defined, measurable and summable in the $p$th degree on the

whole space. For $f(x) \in L_p^{(n)}$ we shall assume that

$$\|f\|_p^{(n)} = \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(x_1, \ldots, x_n)^p \, | \, dx_1 \ldots dx_n \right)^{\frac{1}{p}} \quad (6.3)$$

and in the case where $p = \infty$

$$\|f\|_{\infty}^{(n)} = \sup_{-\infty < x_i < \infty,\ i=1,2,\ldots,n} \text{vrai} \, |f(x_1, \ldots, x_n)|. \quad (6.4)$$

In the ordinary operations of addition of functions and their multiplication by a number, with norms defined by the respective formulae (6.1)–(6.4), spaces $L_{*p}^{(n)}$ and $L_p^{(n)}$ with $1 \leqslant p \leqslant \infty$ are linear normed spaces (here equivalent functions are considered to be identical].

We now note the basic inequalities of S. M. Nikol'skii for integral functions of the exponential type.

Let $\varphi(x_1, \ldots, x_n)$ belong to $L_p^{(n)}$ and for almost every point of $(n-1)$-dimensional space $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ let there be an integral function of degree $\nu$ in the variable $x_i$. Then $\frac{\partial \varphi}{\partial x_i} \in L_p^{(n)}$ and the *generalized Bernstein inequality* is true:

$$\left\| \frac{\partial \varphi}{\partial x_i} \right\|_p^{(n)} \leqslant \nu \, \| \varphi \|_p^{(n)}, \quad 1 \leqslant p \leqslant \infty.$$

Further let $g_{\nu_1, \ldots, \nu_n}(x_1, \ldots, x_n) \in \mathfrak{G}_{\nu_1, \ldots, \nu_n}^{p_1}$ and $1 \leqslant p < p' \leqslant \infty$. Then $g_{\nu_1, \ldots, \nu_n}(x_1, \ldots, x_n) \in \mathfrak{G}_{\nu_1, \ldots, \nu_n}^{p'}$ and we have

$$\| g_{\nu_1 \ldots \nu_n} \|_{p'}^{(n)} \leqslant 2^n \left( \prod_{k=1}^{n} \nu_k \right)^{\frac{1}{p} - \frac{1}{p'}} \| g_{\nu_1 \ldots \nu_n} \|_p^{(n)},$$

where with fixed $n$ and arbitrary $\nu_i \to \infty$ this inequality is exact from the point of view of order.

If again $g_{\nu_1, \ldots, \nu_n}(x_1, \ldots, x_n) \in \mathfrak{G}_{\nu_1, \ldots, \nu_n}^{p}$, $1 \leqslant p \leqslant \infty$ and $1 \leqslant m < n$, then with any fixed $x_{m+1} = x_{m+1}^{(0)}$, $x_{m+2} = x_{m+2}^{(0)}$, $\ldots$, $x_n = x_n^{(0)}$ the function $g_{\nu_1, \ldots, \nu_n}(x_1, \ldots, x_m, x_{m+1}^{(0)}, \ldots, x_n^{(0)})$ is integrable with respect to $(x_1, \ldots, x_m)$ in the $p$th degree and

$$\| g_{\nu_1 \ldots \nu_n} \|_p^{(m)} =$$

$$= \left( \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} | g_{\nu_1 \ldots \nu_n}(x_1, \ldots, x_m, x_{m+1}^{(0)}, \ldots, x_n^{(0)}) |^p \, dx_1 \ldots dx_m \right)^{\frac{1}{p}} \leqslant$$

$$\leqslant 2^n \left( \prod_{k=m+1}^{n} \nu_k \right)^{\frac{1}{p}} \| g_{\nu_1 \ldots \nu_n} \|_p^{(n)},$$

where with fixed $n$ and arbitrary $\nu \to \infty$, it is also exact from the point of view of order. Absolutely similar inequalities were obtained by S. M. Nikol'skii for trigonometric polynomials of class $\mathfrak{T}_{m_1, \ldots, m_n}$. As we have already remarked, these inequalities find natural and varied uses both in the theory of approximation of functions of many variables itself and also in its applications.

As above we now introduce the ideas of best approximations:

$$E^p_{\mathfrak{T}_{m_1 \ldots m_n}}(f) = \inf_{T_{m_1 \ldots m_n} \in \mathfrak{T}_{m_1 \ldots m_n}} \|f - T_{m_1 \ldots m_n}\|^{(n)}_p, \quad f(x) \in L^{(n)}_p,$$

$$E^p_{\mathfrak{G}_{\nu_1 \ldots \nu_n}}(f) = \inf_{g_{\nu_1 \ldots \nu_n} \in \mathfrak{G}^p_{\nu_1 \ldots \nu_n}} \|f - g_{\nu_1 \ldots \nu_n}\|^{(n)}_p, \quad f(x) \in L^{(n)}_p.$$

We shall say that the function $f(x_1, \ldots, x_n)$ has *a generalized partial derivative* of order $k$ with respect to the variable $x_i$ and shall denote it by $\dfrac{\partial^k f}{\partial x_i^k}$, if the function $f(x_1, \ldots, x_n)$ can be transformed on a set of $n$-dimensional measure zero so that after this the function $f(x_1, \ldots, x_n)$ will have a generalized derivative of order $k$ with respect to the variable $x_i$ (in the sense of § 5.2) on any finite segment parallel to the axis of $x_i$ and belonging to the domain of definition of the function $f(x_1, \ldots, x_n)$, for almost all points $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ (in the $(n-1)$-dimensional sense). *Generalized mixed partial derivatives* are defined in a similar way.

Let $1 \leqslant p \leqslant \infty$, $M > 0$, $r > 0$, $r = \bar{r} + \alpha$, $\bar{r}$ is a non-negative whole number, $0 < \alpha \leqslant 1$. The function $f(x)$ is called *a function of class* $H^{(r)}_{px_1}(M)$ if $f(x) \in L^{(n)}_p$ and there exists a generalized partial derivative $\dfrac{\partial^{\bar{r}} f}{\partial x_1^{\bar{r}}} \in L^{(n)}_p$ such that

$$\left\| \frac{\partial^{\bar{r}} f(x_1 + h, x_2, \ldots, x_n)}{\partial x_1^{\bar{r}}} - 2\,\frac{\partial^{\bar{r}} f(x_1, x_2, \ldots, x_n)}{\partial x_1^{\bar{r}}} + \frac{\partial^{\bar{r}} f(x_1 - h, x_2, \ldots, x_n}{\partial x_1^{\bar{r}}} \right\|^{(n)}_p \leqslant M |h|^{\alpha}.$$

In the case where $0 < \alpha < 1$ this is equivalent to the condition

$$\left\| \frac{\partial^{\bar{r}} f(x_1 + h, x_2, \ldots, x_n)}{\partial x_1^{\bar{r}}} - \frac{\partial^{\bar{r}} f(x_1, \ldots, x_r)}{\partial x_1^{\bar{r}}} \right\|^{(n)}_p \leqslant K |h|^{\alpha}$$

for some constant $K > 0$.

In a similar way we define classes $H_{px_i}^{(r)}(M)$, $i = 1, 2, \ldots, n$. The intersection of the classes $H_{px_i}^{(r_i)}(M_i)$ for all $i = 1, 2, \ldots, n$, is represented by $H_p^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n)$. Finally

$$H_p^{(r_1, \ldots, r_n)} = \bigcup_{\substack{M_i > 0 \\ i=1, 2 \ldots, n}} H_p^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n).$$

In exactly the same way in the periodic case we define the classes $H_{*px_i}^{(r_i)}(M)$, $H_{*p}^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n)$ and $H_{*p}^{(r_1, \ldots, r_n)}$.

**6.2. The theory of best approximations of functions of many variables.**

THEOREM 1 (S. M. Nikol'skii). *Let* $f(x) \in H_{*p}^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n)$, *then*

$$E_{\mathfrak{T}_{m_1 \ldots m_n}}^p(f) \leqslant c \sum_{i=1}^n \frac{M_i}{m_i^r i},$$

*where the constant* $c$ *does not depend on* $f$, $M_i$ or $m_i$, $i = 1, 2, \ldots, n$.

THEOREM 1′ (S. M. Nikol'skii). *Let*

$$f(x_1, \ldots, x_n) \in H_p^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n),$$

*then*

$$E_{\mathfrak{G}_{\nu_1 \ldots \nu_n}}^p(f) \leqslant d \sum_{i=1}^n \frac{M_i}{\nu_i^r i},$$

*where the constant* $d$ *does not depend on* $f$, $M_i$ or $\nu_i$, $i = 1, 2, \ldots, n$.

THEOREM 2 (S. M. Nikol'skii). *Let* $f(x_1, \ldots, x) \in L_{*p}^{(n)}$ *and*

$$E_{\mathfrak{T}_{m_1 \ldots m_n}}^p(f) \leqslant \sum_{i=1}^n \frac{K_i}{m_i^r i}, \tag{6.5}$$

*where* $K_i > 0$ *and* $r_i > 0$ *are constants and* $\nu_i$ *form some increasing geometric progression,* $\nu_i = b_i a_i^k$ $(a_i > 1; i = 1, 2, \ldots, n; k = 0, 1, 2, \ldots)$: *also let* $T(x_1, \ldots, x_n)$ *be a trigonometric polynomial not higher than the first degree such that*

$$\left\| f - T \right\|_{*p}^{(n)} \leqslant \sum_{i=1}^n K_i.$$

*Then the polynomial* $T(x_1, \ldots, x_n)$ *(which exists in virtue of condition* (6.5)*) belongs to the class* $H_{*p}^{(r_1, \ldots, r_n)}(M)$, *where*

$$M \leqslant c' \left\| f \right\|_{*p}^{(n)} + d' \sum_{i=1}^n K_i,$$

*and the function* $\varphi(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - T(x_1, \ldots, x_n)$ *belongs to the class* $H_p^{*(r_1, \ldots, r_n)}(M_1, \ldots, M_n)$, *where*

$$M_i \leqslant c_i \sum_{i=1}^{n} K_i,$$

*the constants* $c'$, $d'$ *and* $c_i$ $(i = 1, 2, \ldots, n)$ *not depending on* $f$ *and* $K_i$.

THEOREM 2′ (S. M. Nikol'skii). *Let* $f(x) \in L_n^{(n)}$ *and*

$$E_{\mathfrak{G}_{\nu_1 \ldots \nu_n}}^{p}(f) \leqslant \sum_{i=1}^{n} \frac{K_i}{\nu_i^{r_i}}, \tag{6.6}$$

*where* $K_i > 0$ *and* $r_i > 0$ *are constants, and* $\nu_i$ *form some increasing geometric progression,* $\nu_i = b_i a_i^k$ $(a_i > 1; i = 1, 2, \ldots, n; k = 0, 1, 2, \ldots)$; *also let* $g(x_1, \ldots, x_n)$ *be an integral function of exponential type of degree not higher than the first and such that*

$$\|f - g\|_p^{(n)} \leqslant \sum_{i=1}^{n} K_i.$$

*Then the function* $g(x_1, \ldots, x_n)$ *(which exists in virtue of condition* (6.6)) *belongs to the class* $H_p^{(r_1, \ldots, r_n)}(M)$, *where*

$$M \leqslant c' \|f\|_p^{(n)} + d' \sum_{i=1}^{n} K_i,$$

*and the function* $\varphi(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$ *belongs to class* $H_p^{(r_1, \ldots, r_n)}(M_1, \ldots, M_n)$ *where*

$$M_i \leqslant c_i \sum_{i=1}^{n} K_i,$$

the *constants* $c'$, $d'$ *and* $c_i$ $(i = 1, 2, \ldots, n)$ *not depending on* $f$ *and* $K_i$ $(i = 1, 2, \ldots, n)$.

From Theorems 1 and 2 (1′ and 2′ respectively) it follows in particular that for the fulfilment of the condition $f(x_1, \ldots, x_n) \in H_p^{*(r_1, \ldots, r_n)}$ $(f(x_1, \ldots, x_n) \in H_p^{(r_1, \ldots, r_n)}$ respectively) it is necessary and sufficient that there should exist a constant $K > 0$ such that $E_{\mathfrak{L}_{m_1, \ldots, m_n}}^{p}(f) \leqslant K \sum_{i=1}^{n} \frac{1}{m_i^{r_i}}$ $\left(E_{\mathfrak{G}_{\nu_1, \ldots, \nu_n}}^{p}(f) \leqslant K \sum_{i=1}^{n} \frac{1}{\nu_i^{r_i}}\right.$ respectively$\left.\right)$. Here the given condition remains sufficient in the case where $m_i$ ($\nu_i$ respectively) forms an increasing geometric progression.

The theorems just formulated contain in themselves obviously, as a particular case, the corresponding results of §§ 3.7, 3.9 and 5.2.

Similar questions for the approximation of functions of many variables by algebraic polynomials have been studied much less fully. Little work has been done at the present time on the asymptotic behaviour of the deviations of approximations for various linear methods of approximation of functions of any class. This is also related to other questions of approximations, which we have considered for functions of one variable. Separate information on this question may be found for instance in [1], [3], [8] and [12].

## § 7. THE THEORY OF APPROXIMATIONS IN BANACH SPACES

**7.1. General ideas. Best approximations in Hilbert spaces. The method of least squares for the construction of best approximations.** Let $R$ be a metric space, $E$ one of its subsets, $x_0$ a fixed element of the space $R$; as usual by definition

$$\varrho(x_0, E) = \inf_{x \in E} \varrho(x_0, x)$$

($\varrho(x_0, E)$ is called the *deviation* or *distance* of the element $x_0$ from the set $E$).

If there exists an element $y_0 \in E$ such that

$$\varrho(x_0, E) = \varrho(x_0, y_0),$$

the element $y_0$ is called *the element which realizes the best approximation* of the element $x_0$ on the set $E$.

If the set $E$ is compact and closed then, whatever the element $x_0 \in R$, there always exists an element $y_0 \in E$ which realizes the best approximation of the element $x_0$ on the set $E$.

Let $R$ be a linear normed space. If the set $E \subset R$ is a linear finite-dimensional subspace of the space $R$ then, whatever the element $x_0 \in R$, there always exists an element $y_0 \in E$ which realizes the best approximation of the element $x_0$ on the set $E$; if the space $R$ is strictly normed this element is unique.

We must remember that linear normed space is said to be *strictly normed* if from the relation

$$\|a + b\| = \|a\| + \|b\|$$

it follows that $\lambda a = \mu b$, $\lambda \geqslant 0$, $\mu \geqslant 0$, $\lambda + \mu > 0$.

Spaces $L_p$ with $1 < p < \infty$ and in particular Hilbert space $L_2$ are examples of strictly normed spaces. This ensures in them uni-

queness of best approximations for the respective cases. Spaces $L_1$ and $C$ are not strictly normed. Nevertheless, as we have seen above, in these spaces for a number of important cases there is uniqueness of the element which realizes the best approximation (see also § 7.3).

Now let $H$ be Hilbert space, $E$ one of its subspaces and $x_0 \in H$. The element $y_0 \in E$ is the best approximating element of $x_0$ in the subspace $E$ when and only when, for any $y \in E$,

$$(x_0 - y_0, y) = 0.$$

Hence it follows immediately that the best approximating element in Hilbert space depends linearly on the elements which are being approximated; i.e. if $y_0 \in E$ is the best approximating element in $E$ of the element $x_0 \in H$, and $y_0' \in E$ is the best approximating element in $E$ of the element $x_0' \in H$, then with any numbers $\lambda$ and $\lambda'$ the element $\lambda y_0 + \lambda' y_0'$ is the best approximating element in $E$ of the element $\lambda x_0 + \lambda' x_0'$.

If $E$ is a finite-dimensional subspace of Hilbert space and $\{e_1, \ldots, e_n\}$ its basis, then from what we have said above it follows also that for any element $x_0 \in H$ there exists, and that uniquely, an element $y_0 \in E$ which realizes the best approximation of the given element $x_0$ on the set $E$. If $y_0 = \sum_{i=1}^{n} \lambda_i e_i$ the numbers $\lambda_i$ $(i = 1, 2, \ldots, n)$ are determined from the condition that if $y = \sum_{i=1}^{n} \alpha_i e_i$ the quantity

$$\begin{aligned} J = \|x_0 - y\|^2 &= \left\| x_0 - \sum_{i=1}^{n} \alpha_i e_i \right\|^2 = \\ &= \left( x_0 - \sum_{i=1}^{n} \alpha_i e_i,\ x_0 - \sum_{i=1}^{n} \alpha_i e_i \right) = \\ &= (x_0, x_0) - 2 \sum_{i=1}^{n} \alpha_i (x_0, e_i) + \sum_{\substack{i=1 \\ j=1}}^{n} \alpha_i \alpha_j (e_i, e_j), \end{aligned}$$

considered as a function of $\alpha_1, \ldots, \alpha_n$ must reach a minimum if $\alpha_i = \lambda_i$, $i = 1, \ldots, n$. The written expression containing these elements is a quadratic function. Therefore looking for the best approximation by the given method is often called *the method of least squares*. To find the best approximations in the case where the Hilbert space considered is a subspace of functions with integrable squares this method has already been used by Gauss.

In order to find the quantities $\alpha_i$, $i = 1, 2, \ldots, n$, for which the quantity $J$ is a minimum, it is sufficient to equate to zero the partial derivatives $\dfrac{\partial J}{\partial \alpha_i}$, $i = 1, 2, \ldots, n$. This leads to the system of linear equations

$$\sum_{i=1}^{n} \alpha_i (e_i, e_j), = (x_0, e_j) \quad j = 1, 2, \cdots, n,$$

with the *Gram determinant* $G(e_1, \ldots, e_n)$ not equal to zero in virtue of the linear independence of the system $\{e_1, \ldots, e_n\}$. The value of the deviation $\delta = \|x_0 - y_0\|$ of the element $x_0$ from the element $y_0$, i.e. the best approximation of the element $x_0$ on the set $E$, may in this case be determined by the formula

$$\delta = \sqrt{\frac{G(x_0, e_1, \ldots, e_n)}{G(e_1, \ldots, e_n)}},$$

where the numerator and denominator of the fraction under the root sign are the Gram determinants of the elements standing in the brackets.

**7.2. The connection between the best approximations and the entropy of the sets.** In the present section we shall consider one of the questions of the theory of approximation which is closely connected with the theory, worked out by A. N. Kolmogorov and his school, touching on questions of approximation in Banach spaces and their connection with the entropy of sets [19].

Let $R$ be a separable linear normed space and let

$$X = \{x_0, x_1, \ldots, x_n, \ldots\}, \quad X_n = \{x_0, \ldots, x_n\} \tag{7.1}$$

be a sequence of linearly independent elements of the space $R$ and let

$$E_{X_n}(x) = \inf_{c_1 \ldots c_n} \left\| x - \sum_{k=0}^{n} c_k x_k \right\|_R,$$

where the lower bound of the right-hand side is taken for possible real values of $c_1, \ldots, c_n$.

The system (7.1) is said to be closed if for any $x \in R$

$$\lim_{n \to \infty} E_{X_n}(x) = 0.$$

From now on we shall consider the system $X$ to be fixed and closed. It is obvious that for any element $x \in R$ we have the chain

of inequalities

$$E_{X_0}(x) \geqslant E_{X_1}(x) \geqslant \cdots \geqslant E_{X_n}(x) \geqslant \cdots$$

Conversely, whatever the sequence of numbers

$$\alpha_0 \geqslant \alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant \cdots, \quad \lim_{n\to\infty} \alpha_n = 0,$$

there exists an element $x \in R$ such that

$$E_{X_n}(x) = \alpha_n, \quad n = 0, 1, 2, \ldots$$

Let $A \subset R$. We shall assume that

$$E_{X_n}(A) = \sup_{x\in A} E_{X_n}(x). \tag{7.2}$$

If the set $A$ is compact then $\lim_{n\to\infty} E_{X_n}(A) = 0$.

Let $A$ be a compact set and $\varepsilon > 0$. We shall denote by $N_A(\varepsilon)$ the least number of elements of the space $R$ which form an $\varepsilon$-network for the set $A$. The function

$$I_A(\varepsilon) = \log_2 N_A(\varepsilon)$$

is called *the entropy* of the compact set $A$ (A. N. Kolmogorov). The entropy is obviously a non-increasing function of $\varepsilon$.

Let us assume that

$$\Phi_A(\eta) = \sup_{I_A(\varepsilon)\geqslant\eta} \{\varepsilon\}.$$

The function $\Phi_A(\eta)$ in a sense is the inverse of the entropy $I_A(\varepsilon)$. It appears that the best approximations (7.2) of the compact set $A$ are closely connected with its entropy.

THEOREM 1 (Yu. A. Brudnyi—A. F. Timan). *Let $A$ be a compactum (i.e. a compact metric space), $A \subset R$,*

$$A_X = \{x \in R : \|x\| \leqslant E_{X_0}(A),\ E_{X_k}(x) \leqslant E_{X_k}(A)\}.$$

*Then*

$$E_{X_{n-1}}(A) \leqslant 4\Phi_{A_X}(n+1).$$

In any infinite-dimensional linear normed space there exist compact sets whose entropy increases as quickly as we please as $\varepsilon \to 0$. In this connection it is natural to classify the given sets according to the speed of growth of their entropy as $\varepsilon \to 0$.

We shall represent by $\mathfrak{M}_k^\beta(R)$ the class of sets $A \subset R$ such that

$$\underbrace{\log_2 \log_2 \cdots \log_2}_{k \text{ times}} I_A(\varepsilon) \asymp \left(\log \frac{1}{\varepsilon}\right)^\beta \dagger, \quad k = 0, 1, 2, \ldots, \beta > 0.$$

Note that the set $A$ is of finite dimensions when and only when, for some fixed $n$

$$\inf_X E_{X_n}(A) = 0.$$

The following theorem is in a sense the converse of Theorem 1 and evaluates the best approximations from below.

THEOREM 2. (YU. A. BRUDNYI–A. F. TIMAN). *Let $H$ be Hilbert space,*

$$A \in \mathfrak{M}_k^\beta(H);$$

*then for all sufficiently large $n$:*

*if $k = 0, \beta > 1$,*

$$\inf_X E_{X_n}(A) \geqslant \varrho^{n^{\frac{1}{\beta-1}}}, \quad \varrho < 1;$$

*if $k = 1, 0 < \beta \leqslant 1$,*

$$\inf_X E_{X_n}(A) \geqslant \Phi_A\left[(n \ln n)^{\frac{1}{\beta}}\right];$$

*if $k = 1, \beta > 1$,*

$$\inf_X E_{X_n}(A) \geqslant \varrho^{(\ln n)^{\frac{1}{\beta}}}, \quad \varrho < 1;$$

*and if $k \geqslant 2$,*

$$\inf_X E_{X_n}(A) \geqslant \Phi_A(n^2).$$

**7.3. The theory of best Chebyshev approximations on compacta.** Let $R$ be a compactum. The space of all real functions $f(x)$, which are continuous on $R$, with the natural operations of addition of functions and their multiplication by a number and with norm

$$\|f\|_{C_R} = \max_{x \in R} |f(x)| \tag{7.3}$$

† Following N. Bourbaki we write $f \asymp g$ as $\varepsilon \to 0$ ($f = f(\varepsilon)$, $g = g(\varepsilon)$) if $f = O(g)$ and $g = O(f)$.

forms a linear normed space which we shall represent by $C_R$. Let

$$\Omega = \{\omega_1(x), \ldots, \omega_n(x)\} \tag{7.4}$$

be a system of linearly independent functions, defined and continuous on $R$, and let $\xi_1, \ldots, \xi_n$ be real numbers. We shall call any function of the form

$$P(x) = \sum_{i=1}^{n} \xi_i \omega_i(x)$$

a *generalized polynomial* (with respect to the system $\Omega$). The aggregate of all such generalized polynomials we shall denote by $\mathfrak{P}_\Omega$. For each function $f(x)$ which is continuous on $R$ the quantity

$$\|f - P\|_{C_R} = \max_x |f(x) - P(x)|$$

is called the *deviation* (*distance*) of the function $f(x)$ from the generalized polynomial $P(x) \in \mathfrak{P}_\Omega$ and the quantity

$$E_{\mathfrak{P}_\Omega}^{C_R}(f) = \inf_{P(x) \in \mathfrak{P}_\Omega} \|f - P\|$$

the *best Chebyshev approximation* of the given function by means of the generalized polynomials $P(x) \in \mathfrak{P}_\Omega$.

With the assumptions we have made there always exists a generalized polynomial $P_0(x) \in \mathfrak{P}_\Omega$ such that

$$E_{\mathfrak{P}_\Omega}^{C_R}(f) = \|f - P_0\|_{C_R},$$

i.e. in $\mathfrak{P}_\Omega$ there always exists a generalized polynomial which realizes the best approximation of the given continuous function $f(x)$.

THEOREM 1 (C. de la Vallée-Poussin—E. Ya. Remez). *Suppose that a system of continuous linearly independent functions* (7.4) *and some continuous function $f(x)$ are fixed on a compactum $R$. Then there exists a finite subset $E$ of the compactum $R$, consisting of $r$ points $E = \{x_1, \ldots, x_r\} \subset R$, $1 \leqslant r \leqslant n + 1$, such that the best approximation of the function $f(x)$ by means of generalized polynomials of the system $\mathfrak{P}_\Omega$ on the compactum $R$, in the sense of the norm* (7.3), *coincides with the same best approximation of the function $f(x)$ on the set $E$; here at least one of the generalized polynomials, which least deviates from the function $f(x)$ on $E$, deviates least from this function also on the whole compactum $R$.*

The question of the uniqueness of the generalized best approximating polynomial is resolved by the following theorem.

THEOREM 2 (Haar). *For each function which is continuous on the compactum $R$, in order that there should exist a unique generalized best approximating polynomial $P_0(x) \in \mathfrak{P}_\Omega$, it is necessary and sufficient that each polynomial $P(x) \in \mathfrak{P}_\Omega$, $P(x) \equiv 0$, should have no more than $n - 1$ different zeros on $R$.*

A compactum $R$ which satisfies the conditions of Haar's theorem is called a *Chebyshev set*. In the case where the compactum $R$ lies in $k$-dimensional Euclidean space $E^k$, $k \geqslant 2$, the conditions of Haar's theorem are automatically not fulfilled if $R$ contains inner points. The following theorem explains more fully the structure of Chebyshev sets.

THEOREM 3 (Mèrkh'yuber). *A compact set $R$ which contains not less than two points is a Chebyshev set when and only when $R$ is homeomorphic to some closed set on a circle.*

Here it appears that if on the set $R$ which satisfies Mèrkh'yuber's theorem there exists a unique generalized best approximating polynomial of degree $n$, then with even $n$ the compactum $R$ is homeomorphic to a subset of the segment, and with odd $n$ the compactum $R$ is homeomorphic to a circle or a subset of the segment.

For the Chebyshev approximation of functions which are continuous on a compactum a series of properties of associated approximations of continuous functions is generalized on the segment (for example Chebyshev's Theorem 1 of § 3.5). Information on these results and also on the further development of the theory of Chebyshev approximations may be found in [1], [3], [7], [8], [12] and [15].

**7.4. Some general remarks on the theory of approximation of functions.** We now descibe briefly some questions on the theory of approximation which we have considered above. In functional terms these problems may be described as follows.

We are given a Banach space $B$. For the most part the elements of this space are functions which possess some property or other. In the space $B$ some $n$-dimensional subspace $R^n$ is fixed (e.g. the subset of polynomials of degree not higher than $n - 1$), and the question studied is the approximation in some sense of the elements $f \in B$ by means of elements of the subspace $R^n$. The problems which are studied here are connected first of all with a

group of ideas of the great Russian mathematician P. L. Chebyshev, who was the founder of this branch. Basic questions here are the following:

(1) The question of the existence for each element $f \in B$ of the best approximation in the subspace $R^n$, i.e. of an element $g \in R^n$, for which the quantity $\|f - g\|$ † attains its smallest value.

(2) The question of the uniqueness of the given best approximation for each element $f \in B$.

(3) The question of certain characteristic properties of the elements of the best approximations.

(4) The question of the construction with any predetermined degree of accuracy of the best approximation of the given element $f \in B$ in the given subspace $R^n$.

In the given statement of the problems it appears that the best approximating element always exists, but generally speaking it is not unique. In a number of important and in a sense basic cases, however, it is unique. From the defined point of view the problem of the interpolation of functions by polynomials of a given degree (both algebraic and trigonometric) may also be related to a given group of ideas, since in this case we are also concerned with the approximation of functions by means of elements of a finite dimensional subspace (see § 2).

Problems similar to the above mentioned questions (1)–(4) occur of course in the case where the question arises of the approximation of elements $f \in B$ by means of the elements of some given linear infinite-dimensional subspace $R^\infty \subset B$ (for example the subspace of exponential functions of a given degree).

In all these questions, so long as we have to do with the approximation of elements of space $B$ by means of elements of some linear subspace, the given group of problems is related to the linear problems of the theory of approximation, which were explained in §§ 3.2–3.5 and §§ 7.1–7.3.

Here we must note that the best approximating element itself depends on the approximated elements, generally speaking, in a non-linear manner (e.g. in the space of functions which are continuous on the segment $[a, b]$ with their approximation by means of the subspace of polynomials of degree not higher than $n$). On the other hand in a Hilbert space of functions with a summable

† By the symbol $\|a\|$ we denote as usual the norm of the element $a \in B$.

square† the best approximating element depends linearly on the approximated elements and can be found by means of the so-called method of least squares already studied by Gauss (see § 7.1).

Besides linear problems, in the theory of approximations non-linear problems are also studied which are connected with the approximation of elements $f$ of the space $B$ by means of the elements of some subset $S$ of space $B$, which is not a linear subspace. The founder of this branch is also P. L. Chebyshev, who first considered a problem of similar type, viz. the problem of the best approximation of functions by means of rational functions, the largest possible degrees of whose numerator and denominator are fixed. In non-linear problems the same questions come in which were given above for linear problems, but the assumption here as a rule is much more complicated. Far from there being always in this case a unique best approximating element, there enters here for instance the question of the dimension of the set of best approximations and a series of other new problems arises.

In recent times still more and more the theory of linear and non-linear problems of the theory of approximations is being developed in the given circle of Chebyshev ideas in abstract Banach spaces, i.e. the theory for which the nature of the elements of the space considered is neutral. Unfortunately in view of the limited space in the present reference book we have almost completely left untouched both non-linear problems and approximations in abstract Banach spaces.

Other problems of the theory of approximation of functions are related to the ideas of Weierstrass, Jackson and Bernstein and also touch on approximations of elements of space $B$ by means of the subset $R$, which is presented as the union of some expanding sequence of linear spaces $R_n$, $n = 1, 2, 3, \ldots$, i.e. $R_1 \subset R_2 \subset \subset \cdots \subset R_n \subset \cdots$ and $\bigcup_{n=1}^{\infty} R_n = R$. For example, in the case where $B$ is a Banach space of functions, a set of polynomials (algebraic or trigonometric) of degree not higher than $n$ is often taken as $R_n$, and then the aggregate of all polynomials (algebraic or trigonometric respectively) as the set $R$. The first question which arises here is the question of the principal possibility of as accurate

† And hence in any Hilbert space independently of the nature of its elements.

an approximation as we please of any element $f \in B$ by means of elements $g \in R$ (see § 3.2); and also the question of obtaining these elements $g \in R$ constructively (see § 3.3), if they of course exist. In the case of a positive solution of the question of as accurate an approximation as we please of the element $f \in B$ by means of elements of the set $R$, there arises the question of the speed with which the deviation of the form

$$\omega_n(f) = \inf_{g_n \in R_n} \|f - g_n\|$$

approaches zero as $n \to \infty$, depending on the properties of the element $f$, and also the question of the study of some concrete method of approximation of the elements $f \in B$ by means of elements $g_n \in R_n$, $n = 1, 2, 3, \ldots$ These two questions are studied both as depending on the individual properties of the element $f$ and as depending on some common properties which are characteristic of the element's membership of some class or other. These questions acquire special interest when the space $B$ considered represents a space of functions (§§ 3.7–3.10, § 5 and § 6). Here the study of the properties of approximating functions according to the properties of the approximated functions is also interesting (see Theorem 11 in § 3.7).

Directly connected with the above questions of approximation is one of the most important applications of the theory of approximation of functions which puts forward in a sense the problem which is converse to that just considered (therefore the results in this field are often called converse theorems of the theory of approximations). At first the structural properties of functions are considered according to some given character of the change of value of its deviation from the elements of the space $R_n$ as $n \to \infty$, which are obtained by some method of approximation. Here cases are met where this deviation is considered both on the whole set of definition of the functions and at a point. Secondly, according to some properties of the approximating functions themselves from spaces $R_n$, $n = 1, 2, 3, \ldots$, the properties of the approximated functions are studied. In a number of important cases which we have considered above (see §§ 3, 5 and 6) writers have been successful in studying this question completely in a sense and in describing the structural properties of functions in terms of

approximating functions or their deviation from their approximated function, i.e. in obtaining the necessary and sufficient conditions for the properties of the functions considered in the stated terms. The study of functions according to the properties of some particular approximations or the deviations of functions from the approximating function is generally called *the constructive theory of functions.*

In our reference book we have considered, in spaces, functions of one variable, both periodic and non-periodic. It happens that the periodic case is much simpler than the non-periodic; for instance in the periodic case the problems discussed above have been successfully described in sufficiently full degree only in terms of the best approximations $E_n(\mathfrak{F})$ of a given class of functions $\mathfrak{F}$, belonging to the space of functions $B$.

For a similar description of the properties of non-periodic functions best approximations are found to be insufficient and we must introduce the idea of so-called weighted deviations, i.e. deviations which depend not only on the index $n$ but also on the position of the point on the segment (see § 3.8).

At the present time much attention is being successfully given to the study of various concrete methods of the approximation of functions which are given by means of some algorithm. It appears that, for the approximation of periodic functions, their approximation by means of trigonometric functions, or more accurately by means of linear combinations of sines and cosines of multiple arcs (see § 5.3) is most convenient. Therefore in the periodic case we have considered various methods of the approximation of functions connected first of all with methods of summation of Fourier series (see § 4.1). Here great interest is attached to both theoretical and practical, exact and asymptotic, evaluations of the deviation of the functions $f(x) \in \mathfrak{F}$ from functions $g_n(x) \in R_n$ which are formed by the algorithm of function approximation (e.g. the speed of change of this deviation as $n \to \infty$, its dependence on the properties of functions of class $\mathfrak{F}$, and the method of approximation).

This enables us to evaluate the magnitude of the permissible error in the application of the given method of approximation.

Of particular interest of course are the so-called best linear methods of approximation of functions of some class $\mathfrak{F}$, i.e. those methods which, for a given class of functions, given in accuracy

(or perhaps asymptotically) the same order of approximation as the best approximations. (For more details about this see § 4.4.)

In § 7 of the present chapter we have considered some questions of approximations in Banach spaces, but, as has already been shown above, the questions we have considered here do not reflect fully the now existing direction of development of the theory in this field, and bear only an illustrative character. Because of shortage of space we also have not touched upon the geometric interpretation of the questions of the theory of approximations either in the finite- or infinite-dimensional case.

CHAPTER III

# ALMOST-PERIODIC FUNCTIONS

## § 1. UNIFORM ALMOST-PERIODIC FUNCTIONS ON A STRAIGHT LINE

### 1.1. Various definitions of almost-periodic functions

1. DEFINITION 1. The number $\tau$ is called *an $\varepsilon$-almost-period ($\varepsilon$-displacement)* of the function $f(x)$ $(-\infty < x < \infty)$ if for all $x$

$$|f(x + \tau) - f(x)| < \varepsilon.$$

If $f(x)$ is a periodic function and $p$ its period i.e. $f(x + p) = f(x)$ then obviously $p$ is also an almost-period of $f(x)$ for any $\varepsilon > 0$, and so is any number of the form $np$ $(n = \pm 1, \pm 2, \ldots)$.

DEFINITION 2. The set $E$ of real numbers is said to be *relatively dense* if there exists a number $l > 0$ such that in each interval of the real axis of length $l$ $(\alpha < x < \alpha + l)$ at least one number of the set $E$ can be found.

For example the numbers of the arithmetic progression $np$ $(n = 0, \pm 1, \pm 2, \ldots)$ and numbers of the form $\pm \sqrt{n}$ $(n = 0, 1, 2, \ldots)$ form relatively dense sets. On the other hand numbers of the form $\pm n^2$ do not form a relatively dense set.

DEFINITION 3. A function $f(x)$ which is continuous on the whole real axis is said to be *a uniform almost-periodic* (from now on abbreviated as a.-p.) function, if for each $\varepsilon > 0$ there exists a relatively dense set of $\varepsilon$-almost-periods of the function $f(x)$.

To put it in another way, for each $\varepsilon > 0$ we can find a positive number $l = l(\varepsilon)$ such that, in each interval of length $l$, we can find at least one number $\tau$, for which

$$|f(x + \tau) - f(x)| < \varepsilon \quad (-\infty < x < \infty).$$

It is easy to show that if $f(x)$ is an almost-periodic, but not a periodic, function then if $\varepsilon \to 0$, $l(\varepsilon) \to +\infty$.

Uniform a.-p. functions are also called Bohr a.-p. functions.

**2.** DEFINITION 4. The set of functions $M = \{f(x)\}$, each of which is defined on the whole numerical axis, is said to be *conditionally compact* if from each infinite sequence $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... belonging to the set $M$ we can choose a sequence which converges uniformly on the whole numerical axis.

DEFINITION 5. A function $f(x)$ is said to be *a.-p. (in the Bochner sense)* if the family of functions $\{f(x + h)\}$ $(-\infty < h < \infty)$ is conditionally compact (i.e. if from each infinite sequence $f(x + h_1)$, $f(x + h_2)$ ... we can choose a sequence which converges uniformly on the whole numerical axis).

THEOREM 1 (Bochner). *In the case of continuous functions the classes of Bohr and Bochner a.-p. functions coincide.*

**1.2. The simplest basic properties of almost-periodic functions.** In this paragraph we enumerate the simplest properties of a.-p. functions. Some of these properties follow directly from the definition but some require proofs which are sometimes far from simple.

(1) If $f(x)$ is a uniform a.-p. function then $\alpha f(x)$ and $f(x + c)$ are also uniform a.-p. functions ($c$ is a real number and $\alpha$ a complex).

(2) If $f(x)$ is a uniform a.-p. function then $|f(x)|$ is also a uniform a.-p. function (this follows from the elementary inequality $\|f(x + \tau) - f(x)\| \leqslant |f(x + \tau) - f(x)|$.

(3) If $f(x)$ is a uniform a.-p. function and

$$\inf_{-\infty < x < \infty} |f(x)| = \gamma > 0,$$

then $1/f(x)$ is also a uniform a.-p. function.

There is also a more general statement. We shall denote by $E$ the set of values of the a.-p. function $f(x)$. If the function $F(z)$ is uniformly continuous on the set $E$ then $F[f(x)]$ is a uniform a.-p. function.

(4) A uniform a.-p. function is bounded on the whole numerical axis.

(5) A uniform a.-p. function is uniformly continuous on the whole numerical axis.

(6) The sum and product of a finite number of uniform a.-p. functions are uniform a.-p. functions.

(7) The limit $f(x)$ of a uniform convergent sequence $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... of uniform a.-p. functions is a uniform a.-p. function.

Every trigonometric polynomial

$$S_n(x) = \sum_{k=1}^{n} a_k e^{i\lambda_k x},$$

in which $a_k$ are complex numbers and $\lambda_k$ real numbers is the sum of periodic and consequently almost-periodic functions. On the basis of property (6) $S_n(x)$ is an a.-p. function. From property (7) the uniform limit of a sequence of trigonometric polynomials is a uniform a.-p. function. We also have the converse result, i.e. every uniform a.-p. function is the uniform limit of a sequence of trigonometric polynomials. As distinct from the direct result which, as we have seen, is obtained very easily, the converse result may be obtained only by means of a complicated proof. This converse result is basic in the whole theory of uniform a.-p. functions.

(8) If the derivative $f'(x)$ of a uniform a.-p. function is uniformly continuous then it is also a uniform a.-p. function.

(9) If the indefinite integral

$$F(x) = C + \int_0^x f(x)\,dx$$

of a uniform a.-p. function $f(x)$ is bounded on the whole numerical axis then it is also a uniform a.-p. function (THE BOHL–BOHR THEOREM).

### 1.3. Fourier series

1. THE MEAN VALUE THEOREM. *For every uniform a.-p. function $f(x)$ there exists a mean value*

$$M\{f\} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x)\,dx.$$

AMPLIFIED MEAN VALUE THEOREM. *For every uniform a.-p. function $f(x)$ the limit*

$$\lim_{T\to\infty} \frac{1}{T} \int_a^{a+T} f(x)\,dx = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x+a)\,dx =$$

$$= M\{f(x+a)\} = M\{f(x)\}$$

*exists uniformly with respect to $a$.*

In particular, assuming $a = -T$, we obtain

$$M\{f(x)\} = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{0} f(x)\,dx = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)\,dx.$$

The mean $M\{f\}$ possesses, the following obvious properties:

(1) $M\{cf\} = cM\{f\}$ ($c$ is a constant complex number).

(2) $M\{f + g\} = M\{f\} + M\{g\}$.

(3) $M\{f(x + a)\} = M\{f(x)\}$ ($a$ is an arbitrary real number).

(4) If a sequence of uniform a.-p. functions $f_1(x), f_2(x)$ ..., ..., $f_n(x)$, ... converges uniformly on the whole numerical axis to the function $f(x)$ then

$$\lim_{n\to\infty} M\{f_n\} = M\{\lim_{n\to\infty} f_n\} = M\{f\}.$$

The following property is less obvious and requires proof:

(5) Let $f(x)$ be a uniform a.-p. function. If $f(x) \geqslant 0$ and is not identically equal to zero

$$M\{f(x)\} > 0.$$

2. With any real $\lambda$ the function $e^{-i\lambda x}$ is periodic with period $2\pi/|\lambda|$. Therefore if $f(x)$ is a uniform a.-p. function the product $f(x)e^{-i\lambda x}$ possesses this same property. Consequently the function

$$a(\lambda) = M\{f(x)\, e^{-i\lambda x}\}$$

is defined for all real $\lambda$. The following fact is of fundamental significance:

*The function $a(\lambda)$ may differ from zero very considerably on a denumerable set of values of $\lambda$.*

This statement easily follows from Bessel's inequality for a.-p. functions, which in its turn follows from the property of minimality of the Fourier coefficients.

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be different arbitrary real numbers and $c_1, c_2, \ldots, c_n$ arbitrary complex numbers. Then in virtue of the

obvious transformations,

$$M\left\{\left|f(x)-\sum_{k=1}^{n}c_k e^{i\lambda_k x}\right|^2\right\}=M\{|f(x)|^2\}-\sum_{k=1}^{n}\overline{c_k}M\{f(x)\,e^{-i\lambda_k x}\}-$$

$$-\sum_{k=1}^{n}c_k M\{\overline{f(x)}\,e^{i\lambda_k x}\}+\sum_{k_1=1}^{n}\sum_{k_2=2}^{n}c_{k_1}\overline{c_{k_2}}M\{\exp i\lambda_{k_1}x\times$$

$$\times\exp(-i\lambda_{k_2}x)\}=M\{|f(x)|^2\}+\sum_{k=1}^{n}\overline{c_k}a(\lambda_k)-\sum_{k=1}^{n}c_k\overline{a}(\lambda_k)+$$

$$+\sum_{k=1}^{n}c_k\overline{c_k}=M\{|f(x)|^2\}+\sum_{k=1}^{n}|c_k-a(\lambda_k)|^2-\sum_{k=1}^{n}|a(\lambda_k)|^2.$$

From this identity it follows that the minimum of the left-hand side is attained with $c_k=a(\lambda_k)$. Assuming that $c_k=a(\lambda_k)$ we obtain

$$M\left\{\left|f(x)-\sum_{k=1}^{n}a(\lambda_k)^{i\lambda_k x}\right|^2\right\}=M\{|f(x)|^2\}-\sum_{k=1}^{n}|a(\lambda_k)|^2.$$

Because the left-hand side is non-negative

$$\sum_{k=1}^{n}|a(\lambda_k)|^2\leqslant M\{|f(x)|^2\},\tag{1.1}$$

which is *Bessel's inequality* for a.-p. functions.

From the inequality (1.1) the statement about the function $a(\lambda)$ cited above easily follows. In fact, from inequality (1.1) it follows that for any natural $n$ there is a finite number of values of $\lambda$ for which $\frac{1}{n+1}\leqslant|a(\lambda)|<\frac{1}{n}$. Giving $n$ the values 0, 1, 2, ... we find that there is at most a denumerable set of all $\lambda$ for which $a(\lambda)\neq 0$.

We shall denote those $\lambda$ for which $a(\lambda)\neq 0$ (in arbitrary order) by $\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$, and let $a(\lambda_n)=A_n$.

The numbers $\lambda_n$ are called *Fourier exponents* of the function $f(x)$ and the numbers $A_n$ its *Fourier coefficients*. Thus with every uniform a.-p. function may be associated a Fourier series

$$f(x)\sim\sum_{n}A_n e^{i\lambda_n x},\ A_n=M\{f(x)\,e^{i\lambda_n x}\}.$$

From the inequality (1.1) in the limit it follows that

$$\sum_{n=1}^{\infty}|A_n|^2\leqslant M\{|f(x)|^2\}.\tag{1.2}$$

The inequality (1.2) is also called *Bessel's inequality*. In fact, for every uniform a.-p. function, in the inequality (1.2) the equality sign holds:

$$\sum_{n-1}^{\infty} |A_n|^2 = M\{|f(x)|^2\}. \tag{1.3}$$

The equation (1.3) is called *Parseval's equation* and is one of the basic assertions of the theory of a.-p. functions.

### 1.4. Formal operations on Fourier series

**1.** Let

$$f(x) \sim \sum_n A_n e^{i\lambda_n x} = \sum_\lambda a(\lambda)\, e^{i\lambda x},$$

$$g(x) \sim \sum_n B_n e^{i\mu_n x} = \sum_\lambda b(\lambda)\, e^{i\lambda x}.$$

Then

(1) $kf(x) \sim \sum_n kA_n e^{i\lambda_n x}$ ($k$ is an arbitrary complex number);

(2) $e^{i\lambda x} f(x) \sim \sum_n A_n e^{i(\lambda_n+\lambda)x}$ ($\lambda$ is an arbitrary real number);

(3) $f(x+\alpha) \sim \sum_n A_n e^{i\lambda_n \alpha}\, e^{i\lambda_n x}$ ($\alpha$ is an arbitrary real number);

(4) $\bar{f}(x) \sim \sum_n \overline{A}_n e^{-i\lambda_n x}$;

(5) $f(x) + g(x) \sim \sum_n [a(\lambda) + b(\lambda)]\, e^{i\lambda x}$;

(6) $f(x) \cdot g(x) \sim \sum_n C_n\, e^{i\nu_n x}$,

where

$$C_n = \sum_{\lambda_p+\mu_q=\nu_n} A_p B_q. \tag{1.4}$$

The first five properties are obtained in quite an elementary way. As distinct from them the sixth property may be obtained only by using Parseval's equation (and conversely Parseval's equation may be obtained from equation (1.4)).

**2.** Suppose that we are given a sequence of uniform a.-p. functions

$$f_m(x) \sim \sum_n A_n^{(m)} e^{i\lambda_n^{(m)} x},$$

which converges uniformly for all real $x$ to $f(x)$. Then uniformly for all real $\lambda$

$$\lim_{m\to\infty} M\{f_m(x)\, e^{-i\lambda x}\} = M\{f(x)\, e^{-i\lambda x}\},$$

i.e.

$$f(x) \sim \lim_{m\to\infty} \sum_n A_n^{(m)} e^{i\lambda_n^{(m)} x}.$$

The proof is obvious.

3. If

$$f(x) \sim \sum_n A_n e^{i\lambda_n x}, \tag{1.5}$$

then

$$f'(x) \sim \sum_n i\lambda_n A_n e^{i\lambda_n x}.$$

If

$$F(x) = \int_0^x f(x)\,dx$$

is an a.-p. function then

$$F(x) \sim C + \sum_n \frac{A_n}{i\lambda_n} e^{i\lambda_n x}.$$

4. The function

$$g(x) = M_t\{f(x+t)\,\bar{f}(t)\}$$

is said to be a convolution of the function $f(x)$.

If $f(x)$ is a uniform a.-p. function then $g(x)$ is also such a function. If $f(x)$ has a Fourier series (1.5), then

$$g(x) \sim \sum_n |A_n|^2 e^{i\lambda_n x}.$$

5. If the trigonometric series $\sum_{n=1}^{\infty} a_n e^{i\lambda_n x}$ converges uniformly on the whole numerical axis to the sum $S(x)$ then it is the Fourier series of its own sum.

The proof follows from para. 2 and the properties of orthogonality

$$M\{e^{i\lambda x} e^{-i\mu x}\} = \begin{cases} 0 & \text{if } \lambda \neq \mu, \\ 1 & \text{if } \lambda = \mu. \end{cases}$$

COROLLARY. There exist uniform a.-p. functions whose Fourier exponents form an arbitrary denumerable set.

**1.5. Basic theorems.** The basic theorems of the theory of uniform a.-p. functions are Parseval's equation and the uniqueness theorem (and also the approximation theorem which we shall consider in the following paragraph).

PARSEVAL'S EQUATION (BOHR'S THEOREM). *If* $f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\lambda_n x}$ *is a uniform a.-p. function then*

$$M\{|f(x)|^2\} = \sum_{n=1}^{\infty} |A_n|^2. \tag{1.6}$$

Equation (1.6) is *Parseval's equation.*

THE UNIQUENESS THEOREM. *If for a uniform a.-p. function* $f(x)$ *all the Fourier coefficients are equal to zero, i.e. for all real* $\lambda$

$$M\{f(x)\, e^{-i\lambda x}\} = 0,$$

*the function* $f(x)$ *is identically zero.*

The proofs of these theorems present some difficulties. The equivalence of these theorems is easy to prove in the sense that one of them follows from the other.

(a) From the uniqueness theorem Parseval's equation follows. Let $f(x) \sim \sum_{n} A_n e^{i\lambda_n x}$. Then

$$g(x) = M_t\{f(x+t)\,\overline{f}(t)\} - \sum_{n=1}^{\infty} |A_n|^2\, e^{i\lambda_n x}.$$

Since (in view of Bessel's inequality) $\sum_{n=1}^{\infty} |A_n|^2 < \infty$ the trigonometric series $\sum_{n=1}^{\infty} |A_n|^2 e^{i\lambda_n x}$ converges uniformly and consequently is the Fourier series of its own sum, which we have denoted by $S(x)$. Thus the functions $g(x)$ and $S(x)$ have identical Fourier series. Hence, by the uniqueness theorem, there follows the truth of what we have assumed

$$g(x) = S(x).$$

Assuming that $x = 0$ in this equation we obtain

$$M\{|f(t)|^2\} = \sum_{n=1}^{\infty} |A_n|^2.$$

(b) From Parseval's equation the uniqueness theorem follows. Let all the Fourier coefficients of the function $f(x)$ be equal to zero. Then from Parseval's equation we have

$$M\{|f(x)|^2\} = 0.$$

Hence it follows that $f(x) \equiv 0$ (see § 1.3).

**1.6. The approximation theorem.** For every uniform a.-p. function $f(x)$ and every $\varepsilon > 0$ we can find a trigonometric polynomial

$$P_\varepsilon(x) = \sum_{k=1}^{N} c_k e^{i\nu_k x},$$

which satisfies the inequality

$$|f(x) - P_\varepsilon(x)| < \varepsilon (-\infty < x < \infty).$$

Here we may assume that the exponents $\nu_k$ of the polynomial $P_\varepsilon(x)$ are chosen from the Fourier exponents of the function $f(x)$.

There are various proofs of the approximation theorem. We shall consider one of the methods—that of Bochner. It is first necessary to introduce some new ideas.

DEFINITION 1. A finite or denumerable set of real numbers $\beta_1, \beta_2, \ldots, \beta_n$ is said to be *linearly independent* if there exists no relation of the form

$$r_1\beta_1 + r_2\beta_2 + \cdots + r_n\beta_n = 0 \quad (n = 1, 2, \ldots)$$

with rational numbers $r_1, r_2, \ldots, r_n$ which are not simultaneously equal to zero.

DEFINITION 2. A finite or denumerable set of linearly independent numbers $\beta_1, \beta_2 \ldots$ is called *a basis* of a set of numbers $\Lambda_1, \Lambda_2, \ldots, \Lambda_n, \ldots$, if every number $\Lambda_n$ can be presented in the form of a finite linear combination of the numbers $\beta$, i.e.

$$\Lambda_n = r_1^{(n)}\beta_1 + r_2^{(n)}\beta_2 + \cdots + r_{m_n}^{(n)}\beta_{m_n} \tag{1.7}$$

with rational coefficients.

It is not difficult to show that every denumerable set of real numbers has a basis which is a member of this set.

It is obvious that for one and the same set of numbers there may exist many bases, but in any definite basis the presentation of the number is unique.

If the basis contains a finite number of terms it is said to be *finite* and in the opposite case—*infinite*. If in the presentation (1.7) all $r$ are whole numbers the basis is said to be *integral* and in the opposite case—*rational*.

DEFINITION 3. The function

$$K_n(\beta t) = \sum_{\nu=-n}^{n} \left(1 - \frac{|\nu|}{n}\right) e^{i\nu\beta t} = \frac{\sin^2 \frac{n\beta}{2} t}{n \sin^2 \frac{\beta t}{2}}$$

is called a *Féjer series* ($\beta$ is an arbitrary real number and $n$ an arbitrary natural number).

Let $f(x) \sim \sum_n A_n e^{i\lambda_n x}$ be a uniform a.-p. function. We shall represent by $\beta_1, \beta_2, \ldots$ some basis of Fourier exponents of the function $f(x)$ and by $r, m, n_1, \ldots, n_r$ some positive whole numbers. Let us construct the composite Féjer kernel

$$K_B^{(m)}(t) = K^{(m)}_{\binom{n_1, n_2, \ldots, n_r}{\beta_1, \beta_2, \ldots, \beta_r}}(t) = K_{n_1}\left(\frac{\beta_1 t}{m!}\right) K_{n_2}\left(\frac{\beta_2 t}{m!}\right) \cdots K_{n_r}\left(\frac{\beta_r t}{m!}\right),$$

where $B$ denotes the aggregate of the indices $\binom{n_1, n_2, \ldots, n_r}{\beta_1, \beta_2, \ldots, \beta_r}$.

DEFINITION 4. *The Bochner–Féjer trigonometric polynomial* is the polynomial

$$P_B^{(m)}(x) = M_t\{f(x+t) K_B^{(m)}(t)\} =$$
$$= \sum_{\substack{|\nu_1| \leqslant n_1 \\ \cdots \\ |\nu_r| \leqslant n_r}} k_{n_1, n_2, \ldots, n_r;\ \nu_1, \nu_2, \ldots, \nu_r}\, a\left(\frac{\nu_1}{m!}\beta_1 + \cdots + \frac{\nu_r}{m!}\beta_r\right) \times$$
$$\times \exp i\left(\frac{\nu_1}{m!}\beta_1 + \cdots + \frac{\nu_r}{m!}\beta_r\right) x,$$

where $a\left(\frac{\nu_1}{m!}\beta_1 + \cdots + \frac{\nu_r}{m!}\beta_r\right)$ is the Fourier coefficient of the function $f(x)$ corresponding to the exponent $\frac{\nu_1}{m!}\beta_1 + \cdots + \frac{\nu_r}{m!}\beta_r$, $k_{n_1, \ldots, n_r;\ \nu_1, \ldots, \nu_r}$ being non-negative numbers.

THE APPROXIMATION THEOREM (Bochner). *For every uniform a.-p. function $f(x)$*

$$\lim_{n_1 \to \infty,\ n_2 \to \infty,\ \ldots,\ n_r \to \infty,\ m \to \infty} P_B^{(m)}(x) = f(x),$$

*uniformly on the whole real axis, i.e. for any $\varepsilon > 0$ we can find $r, m, n_1, \ldots, n_r$, such that for all real $x$*

$$|f(x) - P_B^{(m)}(x)| < \varepsilon.$$

**1.7. The convergence of Fourier series for some classes of uniform almost-periodic functions**

**1.** In studying questions of the convergence of Fourier series of a.-p. functions we encounter an additional difficulty in that the Fourier exponents may lie everywhere dense, and therefore

it is not clear in what order to sum the terms of the Fourier series. In the case where the Fourier series converges absolutely the question of the order of its terms is eliminated. First we give two cases of the absolute convergence of Fourier series. Then we shall consider also the case of non-absolute convergence, but on the assumption that the Fourier exponents have isolated limit points.

THEOREM 1. *If the Fourier coefficients of a uniform a.-p. function* $f(x) \sim \sum_n A_n e^{i\lambda_n x}$ *are positive and* $A_n > 0$, *then the series* $\sum_{n=1}^{\infty} A_n$ *is convergent.*

THEOREM 2. *If the Fourier exponents of a uniform a.-p. function* $f(x)$ *are linearly independent the Fourier series for this function converges absolutely*:

$$\sum_{n=1}^{\infty} |A_n| < \infty.$$

**2.** We shall now assume that the Fourier exponents of a uniform a.-p. function have a unique limit point at infinity (all periodic functions in particular are of such a kind).

THEOREM 3. *Let the uniform a.-p. function*

$$f(x) \sim \sum_k A_k e^{i\lambda_k x} \quad (\lambda_{k+1} < \lambda_k,\ \lambda_{-k} = -\lambda_k,\ |A_k| + |A_{-k}| > 0,$$

$\lim_{k\to\infty} \lambda_k = \infty$) *satisfy the Lipschitz condition*

$$\sup_x |f(x+h) - f(x)| < C|h|^{\alpha} \quad (0 < \alpha < 1),$$

*and let*

$$\lim_{k\to\infty} \lambda_k^{-\alpha} \ln \frac{\lambda_{k+1} + \lambda_k}{\lambda_{k+1} - \lambda_k} = 0;$$

*then*

$$f(x) = \lim_{n\to\infty} \sum_{k=-n}^{n} A_k e^{i\lambda_k x}$$

*uniformly on the whole real axis.*

THEOREM 4. *Let* $\lambda_{k+1} - \lambda_k > \alpha > 0$, *where* $\alpha$ *does not depend on* $k$. *Then at every point* $x_0$, *in the neighbourhood of which the function* $f(x)$ *has bounded variation,*

$$\lim_{n\to\infty} \sum_{k=-n}^{n} A_k e^{i\lambda_k x_0} = \frac{1}{2}[f(x_0 + 0) + f(x_0 - 0)].$$

**3. Let us now assume that the Fourier exponents of the function $f(x) \sim \sum_k A_k e^{i\lambda_k x}$ approach zero.**

THEOREM 5. *Let $f(x) \sim \sum_k A_k e^{i\lambda_k x}$ $(\lambda_{k+1} < \lambda_k, \lambda_{-k} = -\lambda_k, |A_k| + |A_{-k}| > 0, \lim_{k\to\infty} \lambda_k = 0)$ be a uniform a.-p. function. Also let there exist a constant C, such that*

$$\left| \int_0^u f(x+u)\, du \right| < C|u|^{1-\alpha} \quad (0 < \alpha \leqslant 1),$$

*and let*

$$\lim_{k\to\infty} \lambda_k^{\alpha} \ln \frac{\lambda_k + \lambda_{k+1}}{\lambda_k - \lambda_{k+1}} = 0.$$

*Then*

$$f(x) = \lim_{n\to\infty} \sum_{k=-n}^{n} A_k e^{i\lambda_k x}$$

*uniformly for all real x.*

**1.8. The relation between the Fourier exponents and almost-periods.** If $f(x)$ is a periodic function with period $p$ its Fourier exponents are multiples of the number $2\pi/p$.

For uniform a.-p. functions there also exists a close connection between the almost-periods and the Fourier exponents of the function, but it is also clearly not so simple as for periodic functions. This connection is established in the following two theorems of Bohr.

THEOREM 1. *Let $f(x) \sim \sum_n A_n e^{i\Lambda_n x}$ be a uniform a.-p. function. For an arbitrary natural number N and a positive number $\delta(<\pi)$ there exists a positive number $\varepsilon = \varepsilon(\delta, N)$ such that all $\varepsilon$-almost-periods $\tau$ of the function $f(x)$ satisfy the system of inequalities* †

$$|\Lambda_k \tau| < \delta \pmod{2\pi}, \quad k = 1, 2, \cdots, N. \tag{1.8}$$

THEOREM 2 (the converse of Theorem 1). *Let $f(x) \sim \sum_n A_n e^{i\Lambda_n x}$ be a uniform a.-p. function. For any $\varepsilon > 0$ we can find a natural*

† Inequality (1.8) means that whole numbers $n_k$ can be found for which the ordinary inequalities

$$|\Lambda_k \tau - 2\pi n_k| < \delta \quad (k = 1, 2, \ldots, N).$$

are satisfied.

*number $N$ and a positive $\delta(<\pi)$ such that every real number $\tau$, which saitsfies the system of inequalities* (1.8), *is an $\varepsilon$-almost-period of the function $f(x)$.*

It is interesting to note that, while the proof of Theorem 1 is quite elementary, that of Theorem 2 is based on the approximation theorem (see [2], Chapter II, § 1).

**1.9. Kronecker's theorem.** Many facts of the theory of uniform a.-p. functions are closely related to the following classical theorem of Kronecker (see in particular the following paragraph, and also the book already mentioned by B. M. Levitan, ch. II).

THEOREM 1 (Kronecker). *Let $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$; $\theta_1, \theta_2, \ldots, \theta_n$ be arbitrary real numbers. In order that the system of inequalities*

$$|\Lambda_k \tau - \theta_k| < \delta (\mathrm{mod}\, 2\pi), \quad k = 1, 2, \ldots, n, \tag{1.9}$$

*may have a completely real solution $\tau$ for any sufficiently small positive number $\delta$, it is necessary and sufficient that from any equation $l_1\Lambda_1 + l_2\Lambda_2 + \cdots + l_n\Lambda_n = 0$, in which $l_1, l_2, \ldots, l_n$ are whole numbers, the following equation*

$$l_1\theta_1 + l_2\theta_2 + \cdots + l_n\theta_n \equiv 0 \ (\mathrm{mod}\, 2\pi) \tag{1.10}$$

*should follow.*

Note two important particular cases of this theorem:

(1) if $\theta_1 = \theta_2 = \cdots = \theta_n = 0$, the system of equations (1.10) is always satisfied and therefore the system of inequalities

$$|\Lambda_k \tau| < \delta \ (\mathrm{mod}\, 2\pi) \tag{1.11}$$

can be solved;†

(2) if the numbers $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ are linearly independent i.e. the equation $l_1\Lambda_1 + l_2\Lambda_2 + \cdots + l_n\Lambda_n = 0$, with whole numbers $l_1, l_2, \ldots, l_n$, is only possible in the case where $l_1 = l_2 = \cdots l_n = 0$, then the condition (1.10) is fulfilled for any $\theta_1, \ldots, \theta_n$ and the system (1.9) can be solved for any $\theta_1, \ldots, \theta_n$ and any $\delta > 0$.

† It is obvious that $\tau = 0$ satisfies system (1.11) with any $\delta > 0$ but the discussion is not concerned with this solution. For the proof of Kronecker's theorem it follows that for any $\delta > 0$ the system (1.9), and then also (1.11), have solutions $\tau$ as large as we please (moreover the numbers $\tau$ form a relatively sense set).

**1.10. Limited periodic functions**

DEFINITION 1. The function $F(x_1, x_2, \ldots, x_m)$ is said to be *limited periodic* if it is the uniform limit of continuous periodic functions $F_k(x_1, x_2, \ldots, x_m)$ $(k = 1, 2, \ldots)$.

If the functions $F_k(x_1, x_2, \ldots, x_m)$ have invariable periods (independent of $k$) with respect to each variable, the limit function will obviously be periodic with the same periods.

DEFINITION 2. The function $F(x_1, x_2, \ldots)$ of a denumerable number of variables, each of which may take on all possible real values, is said to be *limited periodic* if for an arbitrary positive number $\varepsilon$ we can find a positive whole number $n = n(\varepsilon)$ and a continuous periodic function $F_\varepsilon(x_1, x_2, \ldots, x_n)$ such that for arbitrary values of the variables $x_1, x_2, \ldots, x_n, x_{n+1}, \ldots$

$$|F(x_1, x_2, \ldots, x_n, \ldots) - F_\varepsilon(x_1, x_2, \ldots, x_n)| < \varepsilon.$$

$$f(x) = F(x, x, \ldots).$$

THEOREM 1. *For any uniform a.-p. function $f(x)$ there exists a limited periodic function $F(x_1, x_2, \ldots)$ of a finite or denumerable number of variables such that*

$$f(x) = F(x, x, \ldots).$$

The nature of the limited periodic function $F(x_1, x_2, \ldots)$ depends essentially on the kind of bases of the Fourier exponents of the function $f(x)$ (see § 1.6). Thus, for example, if the basis of the Fourier exponents of the function $f(x)$ is a linear combination with integral coefficients of the numbers $\beta_1, \beta_2, \ldots$ then the function $F(x_1, x_2, \ldots)$ is periodic with periods $\frac{2\pi}{\beta_1}, \frac{2\pi}{\beta_2} \ldots$

If the basis is finite (not necessarily integral), i.e. there is a finite number of the numbers $\beta_1, \beta_2, \ldots$, then $F(x_1, x_2, \ldots)$ is a limited periodic function of a finite number of variables.

If the basis is both integral and finite then $F(x_1, x_2, \ldots)$ is a periodic function of a finite number of variables. The corresponding class of uniform a.-p. functions was considered by P. Bohllong before the creation of the general theory of a.-p. functions.

**1.11. The theorem of the argument of a uniform almost-periodic function.** Let $F(t)$ be an arbitrary (complex) continuous, bounded function, defined for all real $t$. Let $\inf_t |F(t)| = k > 0$. Let us consider the function $\varphi(t) = \frac{F(t)}{|F(t)|}$. It is obvious that this

function is also continuous and $|\varphi(t)| = 1$. Let us assume that $\varphi(t) = e^{i\omega(t)}$. If at some point $t_0$ the value of $\omega(t)$ is chosen, then for other values of $t$ we can naturally extend the function $\omega(t)$ as if continuous. It is obvious that in this way the function $\omega(t)$ is defined except for a whole multiple of $2\pi$. If in addition we assume that $-\pi \leqslant \omega(0) < \pi$, the function $\omega(t)$ is defined uniquely.

The function $\omega(t)$ is called the argument of the function $F(t)$ and is denoted by $\arg F(t)$. Various continuous branches of the argument of the function differ by an integral multiple of $2\pi$.

BOHR'S THEOREM. *Let $F(t)$ be a uniform a.-p. function and let $\inf_t |F(t)| = k > 0$. Then $\arg F(t) = ct + \psi(t)$, where $c$ is a constant† and $\psi(t)$ is a uniform a.-p. function. The number $c$ and the Fourier exponents of the function $\psi(t)$ are linear combinations with integral coefficients from the Fourier exponents of the function $F(t)$.*

### 1.12. N-almost-periodic functions

**1.** In this section we consider the class of generalized, continuous, a.-p. functions (which contain the class of uniform a.-p. functions) in whose domain the uniqueness theorem for Fourier series is valid (see § 1.5). In the following section we shall point out the importance of this class of functions for linear differential equations with a.-p. coefficients.

DEFINITION 1. The number $\tau = \tau(\varepsilon, N)$ is said to be an $\varepsilon$, *N-almost-period* of the function $f(x)$ if, for all $|x| < N$,

$$|f(x + \tau) - f(x)| < \varepsilon. \tag{1.12}$$

DEFINITION 2. A continuous function $f(x)$ $(-\infty < x < \infty)$ is said to be *an N-almost-periodic* (N-a.-p. for short) function if we can find a denumerable sequence of real numbers,‡ $\Lambda_1, \Lambda_2, \ldots, \Lambda_n, \ldots$, possessing the property that, whatever positive numbers $\varepsilon$ and $N$ may be, we can find a positive whole number $n = n(\varepsilon, N)$ and a positive number $\delta = \delta(\varepsilon, N)$ such that each real number $\tau$, which satisfies the system of inequalities

$$|\Lambda_k \tau| < \delta \pmod{2\pi} \quad (k = 1, 2, \cdots, n),$$

is an $\varepsilon$, $N$-almost-period of the function $f(x)$, i.e. satisfies the inequality (1.12).

† The constant $c$ is called the mean movement of the a.-p. function $F(t)$.

‡ The numbers $\Lambda_1, \Lambda_2, \ldots, \Lambda_n, \ldots$ clearly depend on the function $f(x)$.

From Theorem 2, § 1.8, it follows that every uniform a.-p. function is an $N$-a.-p. function. That the converse is not true is obvious from the following example.

Let $p(x) = 2 + \cos x\sqrt{2}x$. It is clear that for all $x$, $p(x) > 0$. On the other hand in virtue of Kronecker's theorem (see § 1.9) for an arbitrary $\delta > 0$ we can find a solution of the system of inequalities

$$|x - \pi < \delta \pmod{2\pi},$$

$$\left|\sqrt{2}x - \pi\right| < \delta \pmod{2\pi}.$$

Therefore $\inf p(x) = 0$. Consequently the function $q(x) = \dfrac{1}{p(x)}$ is not bounded and so *cannot be a uniform a.-p.* function (see § 1.2). We shall show, however, that $q(x)$ is an $N$-a.-p. function. Let us choose $N > 0$ and $\varepsilon > 0$ arbitrarily and assume that $k = \min\limits_{N \leqslant x \leqslant N} p(x)$. Next let us choose an arbitrary positive number $\delta < k$ and denote by $\tau = \tau(\delta)$ the almost-period of the function $p(x)$. For $|x| < N$ we have

$$|q(x + \tau) - q(x)| = \frac{|p(x + \tau) - p(x)|}{p(x)\, p(x + \tau)} < \frac{\delta}{k(k - \delta)}.$$

Therefore if $\delta = \min\left(\dfrac{k}{2}, \dfrac{\varepsilon k^2}{2}\right), \dfrac{\delta}{k(k - \delta)}$, and consequently $\tau$ is an $\varepsilon$, $N$-almost-period for $q(x)$ and so $q(x)$ is an $N$-a.-p. function (see § 1.8).

2. THEOREM 1. *The sum, difference and product of N-a.-p. functions is an N-a.-p. function.*

For $N$-a.-p. functions the mean value in the usual sense does not necessarily exist. Let us consider those $N$-a.-p. functions for which

(1) for all real $\lambda$ there exists the limit

$$a(\lambda) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x)\, e^{-i\lambda x}\, dx = M\{f(x)\, e^{-i\lambda x}\},$$

$$\overline{\lim_{T \to \infty}} \frac{1}{2T} \int_{-T}^{T} |f(x)|^2\, dx < \infty. \tag{2}$$

If these conditions are satisfied the function $f(x)$ may be put into a Fourier series

$$f(x) \sim \sum_n' A_n e^{i\lambda_n x}, \quad A_n = M\{f(x)\, e^{-i\lambda_n x}\}.$$

THEOREM 2 (the uniqueness theorem for $N$-a.-p. functions). *If two $N$-a.-p. functions $f(x)$ and $g(x)$ have identical Fourier series these functions are identically equal.*

THEOREM 3 (the approximation theorem for $N$-a.-p. functions). *Let $f(x)$ be an $N$-a.-p. function. For arbitrary $\varepsilon > 0$ and $N > 0$ we can find a trigonometric polynomial $P_{N,\varepsilon}(x) = \sum_{k=1}^{n} a_k e^{i\nu_k x}$ whose Fourier exponents are chosen from the Fourier exponents of the function $f(x)$ and whose coefficients are obtained from the Fourier coefficients of the function $f(x)$, by multiplication by some numbers, and which satisfies the equation*

$$|f(x) - P_{N,\varepsilon}(x)| < \varepsilon \quad \text{for} \quad |x| < N.$$

In particular $P_{N,\varepsilon}(x)$ may be formed by Bochner's method (see § 1.6).

**1.13. Linear differential equations with almost-periodic coefficients.** It is well known that any linear systems of differential equations may be reduced to a system of the form

$$(S_x) \quad \frac{dy_i}{dx} = \sum_{k=1}^{n} f_{ik}(x)\, y_k + g_i(x) \quad (i = 1, 2, \ldots, n).$$

Together with the system $(S_x)$ we shall consider the corresponding homogeneous system

$$(\Sigma_x) \quad \frac{dy_i}{dx} = \sum_{k=1}^{n}{}' f_{ik}(x)\, y_k \quad (i = 1, 2, \ldots, n).$$

Let us consider the case where in the system $(S_x)$ all coefficients $f_{ik}(x)$ and free terms $g_i(x)$ are uniform a.-p. functions. Let $\{h_\nu\}$ be a sequence of real numbers for which there exist the limits

$$\lim_{\nu\to\infty} f_{ik}(x + h_\nu) = f_{ik}^*(x), \quad \lim_{\nu\to\infty} g_i(x + h_\nu) = g_i^*(x)$$
$$(i, k = 1, 2, \ldots, n)$$

(see Bochner's definition, § 1.1, Definition 5).

Along with the systems $(S_x)$ and $(\Sigma_x)$ we may also consider the systems

$$(S_x^*)\quad \frac{dy_i}{dx} = \sum_{k=1}^{n} f_{ik}^*(x)\, y_k + g_i^*(x) \quad (i = 1, 2, \ldots, n),$$

$$(\Sigma_x^*)\quad \frac{dy_i}{dx} = \sum_{k=1}^{n} f_{ik}^*(x)\, y_k \quad (i = 1, 2, \ldots, n).$$

On choosing various sequences $\{h_\nu\}$ we shall obtain families of systems which we shall represent by $H(S_x)$ and $H(\Sigma_x)$.

THEOREM 1. *If a homogeneous system $(\Sigma_x)$ has no bounded solutions (with the exception of the trivial solution $y_k = 0$), then each bounded solution of the system $(S_x)$ consists of N-a.-p. functions.*

THEOREM 2. *If none of the systems $H(\Sigma_x)$ has bounded solutions (except a trivial one) then every bounded solution of the system $(S_x)$ consists of uniform a.-p. functions.*

DEFINITION 1. *The absolute value* of the solution $\{y_i(x)\}$ at the point $x$ is the value

$$\left(\sum_{i=1}^{n} |y_i(x)|^2\right)^{\frac{1}{2}}.$$

THEOREM 3. *If a non-homogeneous system $(S_x)$ has bounded solutions and the homogeneous system $(\Sigma_x)$ has no bounded solution, the absolute value of which can become as small as we please (with the exception of the trivial solution $y_i = 0$), the system $(S_x)$ has at least one solution consisting of N-a.-p. functions.*

THEOREM 4. *If none of the homogeneous systems of the family $H(\Sigma_x)$ has a solution, whose absolute value can become as small as we please (with the exception of a trivial solution), and if the system $(S_x)$ has bounded solutions, then at least one of them consists of uniform a.-p. functions.*

From the last theorem it is easy to deduce that each bounded solution of the differential equation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y = f(x)$$

with constant coefficients $a_1, a_2, \ldots, a_n$ and uniform a.-p. right-hand side $f(x)$, is a uniform a.-p. function.

## § 2. VARIOUS GENERALIZATIONS OF ALMOST-PERIODIC FUNCTIONS

### 2.1. Introductory remarks

DEFINITION 1. *The uniform norm* of the function $f(x)$ is the value

$$\|f\|_u = \sup_{-\infty<x<\infty} |f(x)|.$$

Obviously the uniform norm is finite only for bounded functions.

DEFINITION 2. The quantity

$$D_u(f, g) = \|f - g\|_u \tag{2.1}$$

is called *the distance* between the functions $f(x)$ and $g(x)$.

By the same token the aggregate of bounded functions becomes metric space. Convergence in this space coincides with uniform convergence on the whole numerical axis.

The basic theorem on uniform a.-p. functions may be formulated as follows:

*The set of all uniform a.-p. functions coincides with the closure of the set of all the finite trigonometric sums with respect to the distance* (2.1).

Various extensions of the idea of a.-p. functions to discontinuous (Lebesgue integrable) functions are based on the fact that in the space of functions, which are defined on the whole numerical axis, the distance may be defined in various ways.

### 2.2. The definition and simplest properties of Stepanov almost-periodic functions

DEFINITION 1. The quantity

$$D_{S_l^p}[f(x), g(x)] = \sup_{-\infty<x<\infty} \left[ \frac{1}{l} \int_x^{x+l} |f(x) - g(x)|^p \, dx \right]^{1/p}$$

is called the *S-distance of order* $p \geqslant 1$ *corresponding to the length* $l$ (*$S_l^p$-distance*). The space of functions which are summable in the degree $p$ in each finite interval, with distance as defined above, is called *Stepanov space* (*$S^p$-space*).

*Note.* We can show that $S^p$-distances, corresponding to various $l$, are topologically equivalent. We may therefore limit ourselves to the case $l = 1$. Instead of $S_1^p$ we shall write $S^p$.

THEOREM 1. *$S^p$-spaces are complete, i.e. if*

$$\lim_{\substack{n\to\infty \\ m\to\infty}} D_{S^p}[f_n(x),\ f_m(x)] = 0,$$

*there exists a function $f(x) \in S^p$ such that*

$$\lim_{n\to\infty} D_{S^p}[f(x),\ f_n(x)] = 0.$$

DEFINITION 2. A function $f(x)$ which is measurable and summable in degree $p$ in each finite interval is said to be *$S^p$-almost-periodic* ($S^p$-a.-p. for short) if for any $\varepsilon > 0$ there exists a relatively dense set of numbers $\tau$, satisfying the inequality

$$D_{S_p}[f(x+\tau),\ f(x)] = \sup_{-\infty<x<\infty} \left[\int_x^{x+1} |f(x+\tau) - f(x)|^p\, dx\right]^{1/p} < \varepsilon.$$

The number $\tau$ is called the *$S^p$, $\varepsilon$-almost-period* of the function $f(x)$.

THEOREM 2. *An $S^p$-a.-p. function $f(x)$ is $S^p$-bounded, i.e.*

$$D_{S^p}\{f, 0\} = \sup_{-\infty<x<\infty} \left[\int_x^{x+1} |f(x)|^p\, dx\right]^{1/p} < \infty.$$

THEOREM 3. *An $S^p$-a.-p. function $f(x)$ is $S^p$-uniformly continuous, i.e. for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon)$ such that if $|h| < \delta$, then*

$$D_{S^p}[f(x+h),\ f(x)] < \varepsilon.$$

THEOREM 4. *The sum of two $S^p$-a.-p. functions is an $S^p$-a.-p. function.*

THEOREM 5. *The product of an $S^p$-a.-p. function by an $S^q$-a.-p. function $\left(\frac{1}{p} + \frac{1}{q} = 1\right)$ is an $S$-a.-p. function.*†

THEOREM 6. *If a sequence $\{f(x)\}$ of $S^p$-a.-p. functions converges (in the metric $S^p$) to the function $f(x)$, then the limit function is also $S^p$-a.-p.*

DEFINITION 3. A function $f(x)$ which is summable together with its $p$th degree in each finite interval, is said to be *$S^p$-normal* if the family of functions $\{f(x+h)\}$ ($h$ is an arbitrary real number) is $S^p$-compact, i.e. if from each infinite sequence $f(x+h_1), f(x+h_2), \ldots$ we can choose an $S^p$-convergent sequence.

† Instead of $S^1$ we write $S$.

THEOREM 7. *In order that the function $f(x)$ may be $S^p$-normal, it is necessary and sufficient that it should be $S^p$-a.-p.*

**2.3. The definition and simplest properties of Weyl almost-periodic functions.** In the theory of Stepanov a.-p. functions it is essential that the number $l$, which figures in the description of distance, should be fixed. Conversely for Weyl a.-p. functions it is essential that with decrease of $\varepsilon$ to zero $l$ should increase to $\infty$.

DEFINITION 1. The quantity†

$$D_{W^p}\{f, g\} = \lim_{l\to\infty} D_{S_l^p}\{f, g\} = \lim_{l\to\infty} \sup_{-\infty<x<\infty} \left\{ \frac{1}{l} \int_x^{x+l} |f(x) - g(x)|^p \, dx \right\}^{1/p}$$

is called *W-distance of order p.*

DEFINITION 2. A function $f(x)$, which is integrable in the $p$th degree in each finite interval, is said to be *almost-periodic in the Weyl sense* ($W^p$-a.-*p.* for short) if for every $\varepsilon > 0$ we can find positive numbers $l = l(\varepsilon)$ and $L = L(\varepsilon)$ such that in each interval $(\alpha, \alpha + L)$ of the real axis of length $L$ at least one number $\tau$ can be found which satisfies the inequality

$$D_{S_l^p}[f(x+\tau), f(x)] = \sup_{-\infty<x<\infty} \left\{ \frac{1}{l} \int_x^{x+l} |f(x+\tau) - f(x)|^p \, dx \right\}^{1/p} < \varepsilon.$$

It is obvious that if $\overline{\lim_{\varepsilon\to 0}} l(\varepsilon) < \infty$ then a $W^p$-a.-p. function is an $S^p$-a.-p. function. To be precise it is also obvious that uniform a.-p. functions and $S^p$-a.-p. functions are $W^p$-a.-p. functions.

THEOREM 1. *A $W^p$-a.-p. function $f(x)$ is $W^p$-bounded, i.e.*

$$D_{W^p}\{f(x), 0\} = D_{W^p}\{f(x)\} < \infty.$$

THEOREM 2. *A $W^p$-a.-p. function is $W^p$-uniformly continuous, i.e. for any $\varepsilon > 0$ there exist positive numbers $l = l(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that if $|h| < \delta$ then*

$$D_{S_l^p}\{f(x+h), f(x)\} < \varepsilon.$$

† We may show that this limit always exists.

THEOREM 3. *The sum of two $W^p$-a.-p. functions is a $W^p$-a.-p. function.*

THEOREM 4. *The product of a $W^p$-a.-p. function and a $W^q$-a.-p. function* $\left(\frac{1}{p}+\frac{1}{q}=1\right)$ *is a $W$-a.-p. function.*†

THEOREM 5. *If a sequence of $W_p$-a.-p. functions $\{f_n(x)\}$ is $W^p$-convergent to the function $f(x)$ the latter is $W^p$-a.-p.*

In concluding this section we note that, as distinct from spaces $S^p$, spaces $W^p$ are not complete.

**2.4. The mean value theorem for $W^p$-almost-periodic functions. Parseval's equation for $W^2$- and $S^2$-almost-periodic functions. The approximation theorem.** Since the Fourier coefficients are expressed by definite integrals it is natural to expect that for Stepanov and Weyl a.-p. functions Fourier series also exist. Actually the following theorem is true.

THEOREM 1. *For every $W$-a.-p. function there exists a mean value*

$$M\{f(x)\} = \lim_{T\to\infty} \frac{1}{T}\int_a^{a+T} f(x)\,dx.$$

*The limit of the right-hand side exists uniformly with respect to $a$.*

THEOREM 2. *For any $W^p$-($S^p$-) a.-p. function $f(x)$ the mean value*

$$a(\lambda) = M\{f(x)\,e^{-i\lambda x}\} \qquad (\lambda \text{ is a real number})$$

*differs from zero for at most a denumerable set of values of $\lambda$: $\Lambda_1$, $\Lambda_2$, ..., $\Lambda_n$, ... Thus with every $W^p$-($S^p$-)a.-p. function $f(x)$ may be associated a Fourier series*

$$f(x) \sim \sum_n A_n e^{i\Lambda_n x}, \quad A_n = a(\Lambda_n).$$

THEOREM 3. *For $W^2$-($S^2$-) a.-p. functions Parseval's equation*

$$M\{|f(x)|^2\} = \sum_{n=1}^{\infty} |A_n|^2$$

*is valid.*

THEOREM 4. *For every $W^p$-($S^p$-) a.-p. function $f(x)$ and every $\varepsilon > 0$ we can find a finite trigonometric polynomial $P_\varepsilon$, satisfying the*

† Instead of $W^1$ we write $W$.

*inequality*

$$D_{WP}\{f(x),\ P_\varepsilon(x)\} < \varepsilon \quad [D_{SP}\{f(x),\ P_\varepsilon(x)\} < \varepsilon].$$

*Note.* The polynomial $P_\varepsilon(x)$ may be formed by Bochner's method.

**2.5. Besicovitch almost-periodic functions**

DEFINITION 1. The quantity

$$D_{BP}[f, g] = \left\{ \overline{\lim_{T\to\infty}} \frac{1}{2T} \int_{-T}^{T} |f(x) - g(x)|^p \, dx \right\}^{1/p} = \left\{\overline{M}\{|f - g|^p\}\right\}^{1/p}$$

is called the *$B^p$-distance*. The space of functions, which are summable in the $p$th degree in any finite interval with distance so defined, is called a *$B^p$-space.*

DEFINITION 2. The function $f(x)$ is said to be *almost-periodic in the Besicovitch sense of order $p$* (*a $B^p$-a.-p. function* for short) if there exists a sequence of finite trigonometric sums $P_1(x), P_2(x), \ldots, P_n(x), \ldots$ for which

$$\lim_{n\to\infty} D_{BP}[f(x),\ P_n(x)] = 0.$$

This definition differs in character from the definitions which were given as the basis for the study of $S^p$-$W^p$-a.-p. functions. For these classes of a.-p. functions the definition was based on the generalization of the idea of an almost-period, and what we took as a definition for $B^p$-a.-p. functions was a theorem.

Besicovitch functions, however, can also be defined intrinsically. Such a definition was first given by Besicovitch† but it turned out to be very cumbersome.

A much simpler definition was given by R. Doss.‡ We here give this definition.

In order that the function $f(x)$, which is summable in the $p$th degree ($p \geqslant 1$) in every finite interval, may belong to the class $B^p$-a.-p., it is necessary and sufficient that it should satisfy the following three conditions:

(1) $$\lim_{h\to 0} \overline{M}_x\{|f(x + h) - f(x)|^p\} = 0.$$

† Besicovitch, A. S., *Almost periodic functions.* Cambridge 1932.

‡ Doss, R., On generalized almost periodic functions, *Annals of Maths.* **59**, 3. p. 477 (1954).

(2) For every $\varepsilon > 0$ there exists a relatively dense set of numbers $\tau$ which satisfy the inequality

$$\{\overline{M}_x\{|f(x+\tau) - f(x)|^p\}\}^{1/p} < \varepsilon.$$

(3) For any $a > 0$ there exists a periodic function $f^{(a)}(x)$ of period $a$ and class $L^p$ for which

$$\lim_{n\to\infty} \overline{M}_x\left\{\left|\frac{1}{n}\sum_{k=0}^{n-1} f(x+ka) - f^{(a)}(x)\right|^p\right\} = 0.$$

With every $B^p$-a.-p. function $f(x)$ may be associated a Fourier series

$$f(x) \sim \sum_n A_n e^{i\Lambda_n x}, \quad A_n = M\{f(x)\, e^{-i\Lambda_n x}\}.$$

If two functions $f(x)$ and $g(x)$ have indentical Fourier series their difference $\varphi(x) = f(x) - g(x)$ is a zero function in space $B^p$, i.e.

$$D_{B^p}[\varphi(x), 0] = 0.$$

THEOREM 1. *Spaces $B^p$ ($p \geqslant 1$) are complete.*

COROLLARY (analogue of the Riesz–Fischer theorem). For every trigonometric series

$$\sum_n A_n e^{i\Lambda_n x} \tag{2.2}$$

($A_n$ are complex numbers, $\Lambda_n$ real numbers) for which $\sum_n |A_n|^2 < \infty$ there exists a $B^2$-a.-p. function, whose Fourier series coincides with series (2.2).

**2.6. The idea of almost-periodic functions on groups.** Let $G$ be an *abstract group*, i.e. a set of elements satisfying the following four axioms:

(1) In $G$ multiplication is defined (usually non-commutative) i.e. an operation setting in correspondence with each ordered pair of elements $a, b \in G$ an element $c \in G$:

$$ab = c.$$

(2) The operation of multiplication is associative, i.e. for any three elements $a, b, c \in G$

$$a(bc) = (ab)\, c = abc.$$

(3) In $G$ there exists right-hand identity $e$, i.e. an element such that for any $a \in G$

$$ae = a.$$

(4) For any element $a \in G$ there exists a unique right-hand inverse element $a^{-1}$ such that $a \cdot a^{-1} = e$.

If in addition to these four axioms the condition

$$ab = ba.$$

is also satisfied for any elements $a, b \in G$ the group is said to be *commutative* or *Abelian.*

As an example of an Abelian group we may take the real $n$-dimensional vector space $R^n$, in which the elements of the group are vectors (or points identified with the ends of vectors), and group multiplication coincides with the multiplication of vectors.

As an example of a non-commutative group we may take the case of the group of all rotations of a unit sphere in three-dimensional space.

DEFINITION. The function $f(x)$ ($x \in G$, $f(x)$ takes on complex values) is said to be *almost-periodic on the group* $G$ if the family of functions $f(xa)$, $a \in G$ (or what is equivalent, the family $f(ax)$), is conditionally compact in the sense of uniform convergence on $G$; i.e. from any infinite sequence $f(xa_1)$, $f(xa_2)$, ... we can choose a uniformly convergent subsequence. The theory of a.-p. functions on groups is developed mainly in a similar way to the theory of a.-p. functions on a straight line (a straight line is obviously also a group in which the group operation coincides with ordinary addition). The proof of the following theorems presents a basic difficulty.

THE MEAN VALUE THEOREM. *For each a.-p. function* $f(x)$ ($x \in G$, $G$ *is a group*) *there exists a unique definite value* $M_x\{f(x)\}$ (*the mean value of the function* $f(x)$), *which possesses the following properties.*

(1) $M_x\{\alpha f(x)\} = \alpha M_x\{f(x)\}$ ($\alpha$ *is a complex number*);

(2) $M_x\{f(x) \pm g(x)\} = M_x\{f(x)\} \pm M_x\{g(x)\}$;

(3) $M_x\{1\} = 1$;

(4) *if on the whole group* $G$, $f(x) \geqslant 0$, *then* $M_x\{f(x)\} \geqslant 0$; *if* $f(x) \geqslant 0$ *and at least at one point* $x_0 \in G$, $f(x_0) > 0$, *then* $M_x\{f(x)\} > 0$;

(5) $|M_x\{f(x)\}| \leqslant M_x\{|f(x)|\}$;

(6) $M_x\{\overline{f(x)}\} = \overline{M}_x\{f(x)\}$ (*the bar means the complex-conjugate value*);

(7) $M_x\{f(xa)\} = M_x\{f(ax)\} = M_x\{f(x)\}$ *for an arbitrary element* $a \in G$;

(8) $M_x\{f(x^{-1})\} = M_x\{f(x)\}$.

The existence of a mean value enables us to prove Parseval's equation, the uniqueness theorem for the Fourier series, the approximation theorem, etc. for all a.-p. functions on groups. More details of this can be obtained in the monograph [2], ch. VI.

## § 3. ANALYTIC ALMOST-PERIODIC FUNCTIONS

### 3.1. The definition of analytic almost-periodic functions and their simplest properties

**1.** In the theory of analytic *a.*-p. functions the basic domain of variation of the argument is the strip. In this connection let us agree on a definite terminology.

By $(a, b)$ we shall denote the open interval $a < x < b$; by $[a, b]$ the closed interval $a \leqslant x \leqslant b$. We shall also meet the intervals $(a, b]$, $[a, b)$.

Let $s = \sigma + it$ be a complex variable. The set of all $s$ for which $\sigma = \sigma_0$ will be called *the straight line, the vertical straight line* or simply *the line* $\sigma = \sigma_0$.

The set of all $s$, for which $\sigma$ belongs to the interval $(a, b)$ we shall call *the strip* $(a, b)$. We shall also meet the strips $[a, b]$, $[a, b)$ and $(a, b]$.

2. DEFINITION 1. A real number $\tau$ is said to be an *$\varepsilon$-almost-period* of the analytic function $f(s)$ in the strip $(\alpha, \beta)$, if at all points of the strip $(\alpha, \beta)$

$$|f(s + i\tau) - f(s)| < \varepsilon.$$

We shall represent the set of all $\varepsilon$-almost-periods of the function $f(s)$ (in the strip considered) by $E\{\varepsilon; f(s)\}$.

DEFINITION 2. The function $f(s)$ is said to be *a.-p. in the strip* $(\alpha, \beta)$ if for each $\varepsilon > 0$ the set $E\{\varepsilon; f(s)\}$ is relatively dense.

We may similarly define almost-periodicity in the strip $[\alpha, \beta]$†

† Here we require that $f(s)$ shall be regular in the strip $(\alpha, \beta)$ and continuous in the strip $[\alpha, \beta]$.

THEOREM 1. *A function $f(s)$ which is a.-p. in the strip $[\alpha, \beta]$ is bounded in this strip.*

THEOREM 2. *Let $f(s)$ be regular in the strip $(\alpha, \beta)$ and bounded in any strip $[\alpha_1, \beta_1]$, $(\alpha < \alpha_1 < \beta_1 < \beta)$. Let us assume that on some straight line $\sigma = \sigma_0 (\alpha < \sigma_0 < \beta)$ $f(s)$ is a uniform a.-p. function. Then $f(s)$ is an a.-p. function in the strip $[\alpha_1, \beta_1]$*

*Note.* This theorem is not true if the condition of boundedness in any strip $[\alpha_1, \beta_1]$ is not fulfilled.

THEOREM 3. *The sum and product of a.-p. functions in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$* possess the same properties.

COROLLARY. The polynomial $\sum_{n=1}^{N} a_n e^{\lambda_n s}$ ($\lambda_n$ are real numbers, $a_n$ complex numbers) is an a.-p. function in the strip $(-\infty, \infty)$.

THEOREM 4. *The uniform limit of the sequence of analytic a.-p. functions $\{f_n(s)\}$ in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ is an a.-p.-function in the strip $[\alpha_1, \beta_1]$.*

COROLLARY. The sum of the series $\sum_{n=1}^{\infty} a_n e^{\lambda_n s} (\lambda_n$ is real) which is uniformly convergent in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ is an a.-p. function in this strip.

THEOREM 5. *The derivative of a function which is a.-p. in the strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ is also an a.-p. function in this strip.*

THEOREM 6. *If the indefinite integral $F(s)$ of a function $f(s)$, which is a.-p. in the strip $[\alpha_1, \beta_1]$, is bounded in this strip, then it is an a.-p. function in this same strip.*

THEOREM 7. *Let $F(s)$ be a function which is a.-p. in the strip $(\alpha, \beta)$. We shall denote by $G$ the set of all values of the function $f(s)$ on the line $\sigma_0 (\alpha < \sigma_0 < \beta)$. Whatever the value of $\delta > 0$ the function $f(s)$ assumes all the values of the derived set $G'$ in the strip $(\sigma_0 - \delta, \sigma_0 + \delta)$.*

COROLLARY 1. If the function $f(s)$ is a.-p. in the strip $(\alpha, \beta)$ and does not vanish in this strip, then $\inf |f(s)| > 0$ in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$.

COROLLARY 2. If $f(s)$ is an a.-p. function in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ and $f(s) \neq 0$ in the strip $(a, \beta)$ then $g(s) = \dfrac{1}{f(s)}$ is an a.-p. function in any strip $[\alpha_1, \beta_1]$.

THEOREM 8. *Let $f(s)$ and $g(s)$ be two analytic a.-p. functions in the strip $(\alpha, \beta)$. If their ratio $h(s) = \dfrac{f(s)}{g(s)}$ is bounded in the strip, then $h(s)$ is an a.-p. function in any strip $(\alpha_1, \beta_1)$ $(\alpha < \alpha_1 < \beta_1 < \beta)$.*

### 3.2. Dirichlet series

**1.** To every analytic a.-p. function there corresponds an infinite series which is called a Dirichlet series. This series is similar to the Fourier series of uniform a.-p. functions of a real variable.

THEOREM 1. *Let* $f(s) = f(\sigma + it)$ *be an analytic a.-p. function in the strip* $(\alpha, \beta)$. *We shall consider the function* $f(\sigma + it)$ *with fixed* $\sigma$. *The Fourier series of this function has the form*

$$f(s) = f(\sigma + it) \sim \sum_n A_n e^{\Lambda_n \sigma} e^{i\Lambda_n t} = \sum_n A_n e^{\Lambda_n s}, \qquad (3.1)$$

*where the numbers* $A_n$ *and* $\Lambda_n$ *do not depend on* $s$.

The series (3.1) is called the *Dirichlet series* for $f(s)$. For $A_n$ we have the formula

$$A_n = M_t\{f(\sigma + it)\, e^{-\Lambda_n \sigma} e^{-i\Lambda_n t}\} \quad (\alpha < \sigma < \beta).$$

THEOREM 2. *If* $f(s) = \sum_{n=1}^{\infty} A_n e^{\Lambda_n s}$, *where the series converges uniformly in some strip, then it is a Dirichlet series for the sum* $f(s)$.

THEOREM 3 (the uniqueness theorem for a Dirichlet series). *If two functions which are analytic and a.-p. in one and the same strip have the same Fourier series then they are identical in this strip.*

THEOREM 4 (Parseval's equation). *Let* $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ *be an a.-p. function in the strip* $(\alpha, \beta)$. *For every* $\sigma$ *from the strip* $(\alpha, \beta)$

$$M_t\{|f(\sigma + it)|^2\} = \sum_{n=1}^{\infty} |A_n|^2 e^{2\Lambda_n \sigma}.$$

THEOREM 5 (the approximation theorem). *Let* $f(s)$ *be an a.-p. function in the strip* $(\alpha, \beta)$. *The sequence of the Bochner–Fejer sums for the Dirichlet series of the function* $f(s)$ *converges uniformly to* $f(s)$ *in any inner strip.*

**2.** In view of Theorem 2, § 3.1, for the almost-periodicity of a function $f(s)$ in any strip $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ it is sufficient that it should be bounded in this strip and almost-periodic on at least one straight line inside the strip. We can find another interesting sign of almost-periodicity of the function in the strip, into which the boundedness of the function does not enter.

THEOREM 6. *Let* $F(s)$ *satisfy the following conditions: it is* (1) *regular in the strip* $(\alpha, \beta)$, (2) *continuous in the strip* $[\alpha, \beta]$ *and* (3) *on the lines* $\sigma = \alpha$ and $\sigma = \beta$, $F(s)$ *is a uniform a.-p. function of the variable* $t$ *with Fourier series* $\sum_n A_n e^{\Lambda_n \sigma} e^{i\Lambda_n t} (\sigma = \alpha, \sigma = \beta)$.

*If these conditions are satisfied $F(s)$ is an a.-p. function in the strip $[\alpha, \beta]$ with Dirichlet series*

$$F(s) \sim \sum_n A_n e^{\Lambda_n s}.$$

THEOREM 7. *Suppose we are given a series*

$$\sum_n A_n e^{\Lambda_n s} \tag{3.2}$$

*with negative Dirichlet exponents ($\Lambda_n < 0$). We shall assume that for some $\sigma = \alpha$ the series (3.2) is the Fourier series of the a.-p. function $f_\alpha(t)$. Then there exists a function which is a.-p. in the strip $(\alpha, +\infty)$ and continuous in the strip $[\alpha, +\infty]$ and which coincides with the function $f_\alpha(t)$ on the line $\sigma = \alpha$, has as its Dirichlet series the series (3.2) and approaches zero uniformly with respect to $t$ as $\sigma \to +\infty$.*

THEOREM 8. *Let $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ be an a.-p. function in the strip $(\alpha, \beta)$ where all $\Lambda_n < 0$. Then $f(s)$ is an a.-p. function in the strip $(\alpha, +\infty)$ and approaches zero uniformly with respect to $t$.*

*Note.* Similar theorems hold for the series with exclusively positive Dirichlet exponents.

COROLLARY. If the Dirichlet exponents of an a.-p. function are bounded in some strip $(\alpha, \beta)$ then this function is an integral analytic function.

**3.** By virtue of Theorem 8 an analytic a.-p. function with negative Dirichlet exponents uniformly approaches zero as $\sigma \to +\infty$ and, consequently, is bounded (as $\sigma \to +\infty$). The converse also holds.

THEOREM 9. *Let the function $f(s)$ be a.-p. in the strip $(\alpha, +\infty)$ and uniformly bounded in the half-plane $\sigma > \alpha$. Then all the Dirichlet exponents of the function $f(s)$ are non-positive.*

THEOREM 10. *Let $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ be an a.-p. function in any strip $[\alpha, \beta]$. If the indefinite integral $F(s)$ of the function $f(s)$ is bounded (and consequently is an a.-p. function) in the strip $[\alpha, \beta]$ then the series*

$$\sum_{\Lambda_n < 0} A_n e^{\Lambda_n s}, \quad \sum_{\Lambda_n > 0} A_n e^{\Lambda_n s}$$

*are Dirichlet series of the two functions $f_1(s)$ and $f_2(s)$, a.-p. in any strip $[\alpha_1, +\infty)$, $\alpha_1 > \alpha$, and $(-\infty, \beta_1]$, $\beta_1 < \beta$, respectively.*

**3.3. The convergence of the Dirichlet series for analytic almost-periodic functions.** In the following theorem some tests for the absolute convergence of Dirichlet series are given.

THEOREM 1. *Let* $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ *be an a.-p. function in any strip* $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$.

*If one of the following conditions is satisfied*:

(1) $\Lambda_n$ *are linearly independent,*

(2) $A_n > 0$,

(3) *for all* $\delta > 0$ *the series* $\sum_n e^{-|\Lambda_n|\delta}$ *converges, then the series* $\sum_n A_n e^{\Lambda_n s}$ *converges absolutely in the strip* $(\alpha, \beta)$.

Case (2) enables us to prove the following theorem for analytic a.-p. functions.

THEOREM 2. *Suppose that all the Dirichlet coefficients* $A_n$ *of the a.-p. function* $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ *are positive. If* $(\alpha, \beta)$ *is the maximum strip of the almost-periodicity of the function* $f(s)$ *then the points* $s = \alpha$ *and* $s = \beta$ *are singular for* $f(s)$.

**3.4. The behaviour of almost-periodic function when $\sigma = +\infty$ (analogues of the Weierstrass—Sokhotskii and Picard theorems)**

**1.** THEOREM 1 (analogue of the Weierstrass–Sokhotskii theorem). *Let* $f(s)$ *be an a.-p. function in the strip* $[\alpha, +\infty)$. *We can state only three possibilites about the behaviour of the function* $f(s)$ *in the neighbourhood of the point* $\sigma = +\infty$.

(a) $f(s)$ *approaches a finite limit as* $\sigma \to +\infty$ *and does so uniformly with respect to* $t$.

(b) $|f(s)| \to +\infty$ *as* $\sigma \to +\infty$ *uniformly with respect to* $t$.

(c) *In each half-plane* $\sigma > \alpha_1 > \alpha$ $f(s)$ *takes on values as near as we please to each complex number.*

We shall say that the function $f(s)$ belongs to class (a), (b) or (c) depending on whichever case is relevant.

The following theorem makes it possible for any Dirichlet series of the function $f(s)$ for us to predetermine the class to which it belongs.

THEOREM 2. *Let* $f(s) \sim \sum_n A_n e^{\Lambda_n s}$ *be an a.-p. function in the strip* $(\alpha, \beta)$ *with any* $\beta > \alpha$. *The function* $f(s)$ *belongs*:

(1) *to class* (*a*) *if all Dirichlet exponents are non-positive,*

(2) *to class* (*b*) *if there are positive Dirichlet exponents and among them a greatest, and*

(3) *to class* (*c*) *if there are positive Dirichlet exponents, but among them no greatest.*

**2.** Let us now consider theorems of the same type as Picard's.

THEOREM 3. *Let* $f(s)$ *be an a.-p. function in the strip* $(\alpha, +\infty)$ *and let it belong to class* (*C*). *Then in each half-plane* $(\alpha_1, +\infty)$ $(\alpha_1 > \alpha)$ *the function* $f(s)$ *assumes all values, with the exception perhaps of one* (*i.e.* $f(s)$ *has no more than one exclusive Picard value*).

THEOREM 4. *If the positive Dirichlet exponents of the function* $f(s)$ *have a finite upper bound* $\Lambda$, *while the number* $\Lambda$ *itself is not a Dirichlet exponent, the function* $f(s)$ *has absolutely no exclusive Picard values.*

### 3.5. Harmonic almost-periodic functions

**1.** Let $f(s) = u(\sigma, t) + iv(\sigma, t)$ $(s = \sigma + it)$ be an a.-p. function in the strip $(\alpha, \beta)$. It is known that $u(\sigma, t)$ and $v(\sigma, t)$ are harmonic functions in the same strip. In this paragraph we consider a.-p. functions which are harmonic in the same strip.

DEFINITION 1. A function $u(\sigma, t)$ which is harmonic in the strip $[\sigma_1, \sigma_2]$ is said to be an a.-p. function in the strip, if for every $\varepsilon > 0$ there exists a relatively dense set of $\varepsilon$-almost-periods, i.e. of numbers $\tau$ for which

$$|u(\sigma, t + \tau) - u(\sigma, t)| < \varepsilon \quad (-\infty < t < \infty;\ \sigma_1 \leqslant \sigma \leqslant \sigma_2).$$

THEOREM 1. *Let* $u(\sigma, t)$ *be a harmonic and bounded function in the strip* $[\sigma_1, \sigma_2]$. *If on the lines* $\sigma = \sigma_1$ and $\sigma = \sigma_2$ $u(\sigma, t)$ *is a uniform a.-p. function, then it is a harmonic a.-p. function in the strip* $[\sigma_1, \sigma_2]$.

THEOREM 2. *The sum of a finite number of harmonic a.-p. functions in the strip* $(\sigma_1, \sigma_2)$ *is a harmonic a.-p. function in the same strip. The limit of a uniformly convergent sequence of harmonic a.-p. functions in the strip* $[\sigma_1, \sigma_2]$ *is also a harmonic a.-p. function in the same strip.*

THEOREM 3. *The partial derivatives of all orders of a function which is harmonic a.-p. in the strip* $[\sigma_1, \sigma_2]$ *are also harmonic a.-p. functions in the same strip.*

2. To every harmonic a.-p. function $u(\sigma, t)$ may be related a Fourier series of the form

$$u(\sigma, t) \sim k\sigma + l + \sum_n (A_n^+ e^{\Lambda_n \sigma} + A_n^- e^{-\Lambda_n \sigma}) \cos \Lambda_n t + \\ + \sum_n (B_n^+ e^{\Lambda_n \sigma} + B_n^- e^{-\Lambda_n \sigma}) \sin \Lambda_n t, \tag{3.3}$$

where $k$, $l$, $A_n^+$, $A_n^-$, $B_n^+$, $B_n^-$ do not depend on $\sigma$ and $t$.

THEOREM 4 (the uniqueness theorem for Fourier series). *Let $u_1(\sigma, t)$ and $u_2(\sigma, t)$ be harmonic and a.-p. functions in the strip $[\sigma_1, \sigma_2]$. If the Fourier series for the functions coincide then in this strip $u_1 \equiv u_2$.*

THEOREM 5 (the approximation theorem). *The Bochner–Féjer sums for a harmonic a.-p. function $u(\sigma, t)$ in the strip $(\sigma_1, \sigma_2)$ converge uniformly to $u(\sigma, t)$ in every inner strip $[\sigma_1', \sigma_2'] (\sigma_1 < \sigma_1' < \sigma_2' < \sigma_2)$.*

### 3.6. The mean movement and density of values of analytic almost-periodic functions

1. An analytic function may vanish only at isolated points. Therefore for every analytic function (independent of whether it vanishes or not) the idea of the argument of the function may be reasonably introduced.

Let the function $f(s)$ $(s = \sigma + it)$ be analytic in some domain $G$ and differ in this domain from identical zero. At all points of the domain $G$, excluding the zeros of $f(s)$, the function arg $f(s)$ can be determined except for a multiple of $2\pi$, by the formula $f(s) = |f(s)|\, e^{i \arg f(s)}$.

Let us assume that $L$ is a straight line (or segment of a straight line) from $G$. We shall in future assume that the straight line $L$ is orientated, i. e. on $L$ the positive and negative directions differ.

We shall define the function $\arg^- f(s)$ $(s \in L)$ by taking an arbitrary branch of the argument of the function $f(x)$ and producing it further at all points where $f(s) \neq 0$ by continuity. If $s$ passes through a zero of multiplicity $p$ of the function $f(s)$ in the positive direction, then to the function $\arg^- f(s)$ can be ascribed a jump $-p\pi$. The function $\arg^- f(s)$ is called the *left-hand argument*.

We define similarly the right-hand argument, $\arg^+ f(s)$, on $L$ as the arbitrary branch of the argument of $f(s)$ which is continuous everywhere, with the exception of zeros of $f(s)$ on $L$ and with a jump $+p\pi$, if $s$ passes through a zero of $f(s)$ of multiplicity $p$ in the positive direction.

At points of discontinuity of the functions just defined we shall ascribe to the function a mean arithmetic value. Both functions $\arg^- f(s)$ and $\arg^+ f(s)$ are thus defined at all points $s$ on $L$, clearly with accuracy to within a multiple of $2\pi$. If $f(s)$ has no zeros on $L$ then each of the functions coincides with some continuous branch of $\arg f(s)$ on $L$. Let $s_1$ and $s_2$ be two points on $L$ situated in the positive direction of the straight line $L$. Then the differences

$$\arg^- f(s_2) - \arg^- f(s_1), \quad \arg^+ f(s_2) - \arg^+ f(s_1)$$

do not depend on the choice of the branch of the argument and are called the *change (variation)* of $\arg f(s)$ *within the limits* $s_1$ *and* $s_2$ *in the left-hand (or right-hand) direction.* It is obvious that

$$\arg^- f(s_2) - \arg^- f(s_1) \leqslant \arg^+ f(s_2) - \arg^+ f(s_1).$$

2. Let $f(s)$ be an analytical a.-p. function in the strip $(\alpha, \beta)$. Let $\sigma$ belong to the interval $(\alpha, \beta)$. We shall consider the straight line $L$: $s = \sigma + it$, as orientated according to the increase of $t$.

We now introduce into the discussion the four numbers:

$$\underline{C}^{\pm}(\sigma) = \varliminf_{(\delta-\gamma)\to\infty} \frac{\arg^{\pm} f(\sigma + i\delta) - \arg^{\pm} f(\sigma + i\gamma)}{(\delta - \gamma)},$$

$$\overline{C}^{\pm}(\sigma) = \varlimsup_{(\delta-\gamma)\to\infty} \frac{\arg^{\pm} f(\sigma + i\delta) - \arg^{\pm} f(\sigma + i\gamma)}{(\delta - \gamma)}.$$

$\underline{C}^-(\sigma)$ is called the *lower left-hand mean movement* of the function $f(s)$ on the straight line $s = \sigma + it$ $(-\infty < t < \infty)$, and $\overline{C}^-(\sigma)$ the *upper left-hand mean movement*, etc. These numbers are finite (if $f(s)$ is a.-p. in the strip).

If $\underline{C}^-(\sigma) = \overline{C}^-(\sigma) = C^-(\sigma)$, $\underline{C}^+(\sigma) = \overline{C}^+(\sigma) = C^+(\sigma)$, the numbers $C^-(\sigma)$ and $C^+(\sigma)$ are called *the left-hand and right-hand mean movements* respectively of $f(s)$ on the straight line $\sigma$.

Let $\alpha < \sigma_1 < \sigma_2 < \beta$ and $-\infty < \gamma < \delta < +\infty$. We shall denote by $N(\sigma_1, \sigma_2; \gamma, \delta)$ the number of zeros of the function $f(s)$ in the rectangle $\sigma_1 < \sigma < \sigma_2$, $\gamma < t < \delta$. The numbers

$$\underline{H}(\sigma_1, \sigma_2) = \varliminf_{(\delta-\gamma)\to\infty} \frac{N(\sigma_1, \sigma_2; \gamma, \delta)}{(\delta - \gamma)},$$

$$\overline{H}(\sigma_1, \sigma_2) = \varlimsup_{(\delta-\gamma)\to\infty} \frac{N(\sigma_1, \sigma_2; \gamma, \delta)}{(\delta - \gamma)}$$

are called the *lower* and *upper density of zeros* respectively of the function $f(s)$ in the strip $(\sigma_1, \sigma_2)$. For an a.-p. function the numbers are finite.

If $\underline{H} = \overline{H} = H$ the number $H = H(\sigma_1, \sigma_2)$ is called the *density of zeros* of $f(s)$ in the strip $(\sigma_1, \sigma_2)$.

In the following theorem the connection between the mean movements and the density of zeros of an a.-p. function is established.

THEOREM 1. *Let $f(s)$ be an analytic a.-p. function in the strip $(\alpha, \beta)$, which differs from identical zero. The mean movements and densities of zeros in an arbitrary strip $(\sigma_1, \sigma_2)$ $(\alpha < \sigma_1 < \sigma_2 < \beta)$ are connected by the following inequalities:*

$$\frac{1}{2\pi}[\underline{C}^-(\sigma_2) - \overline{C}^+(\sigma_1)] \leqslant \underline{H}(\sigma_1, \sigma_2) \leqslant \overline{H}(\sigma_1, \sigma_2) \leqslant \frac{1}{2\pi}[\overline{C}^-(\sigma_2) - \underline{C}^+(\sigma_1)].$$

COROLLARIES:

(1) If $C^+(\sigma)$ exists, then for each $\sigma_2 > \sigma_1$

$$\underline{H}(\sigma_1, \sigma_2) = \frac{1}{2\pi}[\underline{C}^-(\sigma_2) - C^+(\sigma_1)],$$

$$\overline{H}(\sigma_1, \sigma_2) = \frac{1}{2\pi}[\overline{C}^-(\sigma_2) - C^+(\sigma_1)].$$

(2) If $C^-(\sigma_2)$ exists, then for each $\sigma_1 < \sigma_2$

$$\underline{H}(\sigma_1, \sigma_2) = \frac{1}{2\pi}[C^-(\sigma_2) - \overline{C}^+(\sigma_1)],$$

$$\overline{H}(\sigma_1, \sigma_2) = \frac{1}{2\pi}[C^-(\sigma_2) - \underline{C}^+(\sigma_1)].$$

(3) If there exist two of the three quantities $C^+(\sigma_1)$, $C^-(\sigma_2)$, $H(\sigma_1, \sigma_2)$, the third value also exists, where

$$H(\sigma_1, \sigma_2) = \frac{1}{2\pi}[C^-(\sigma_2) - C^+(\sigma_1)].$$

**3. JENSEN'S FUNCTION.**

THEOREM 2. *For every a.-p. function $f(s)$ in the strip $(\alpha, \beta)$ which is not identically zero, there exists uniformly with respect to $\sigma$ in the*

*interval* $[\alpha_1, \beta_1]$ $(\alpha < \alpha_1 < \beta_1 < \beta)$ *a mean value*

$$\varphi(\sigma) = M_t\{\ln |f(\sigma + it)|\},$$

*i.e. the function*

$$\varphi(\sigma; \gamma, \delta) = \frac{1}{(\delta - \gamma)} \int_{\gamma}^{\delta} \ln |f(\sigma + it)|\, dt$$

*approaches the limit* $\varphi(\sigma)$ *uniformly with respect to* $\sigma$ *in* $[\alpha_1, \beta_1]$ *as* $(\delta - \gamma) \to +\infty$.

The function $\varphi(\sigma)$ may also be defined in the following way:

Let $m > 0$ be chosen arbitrarily. Let us assume that $|f(s)|_m = \max\{|f(s)|, m\}$

$$\varphi_m(\sigma) = M_t\{\ln |f(\sigma + it)|_m\}.$$

As $m \to 0$, $\varphi_m(\sigma)$ converges uniformly in $[\alpha_1, \beta_1]$ to $\varphi(\sigma)$. The function $\varphi(\sigma)$ is called *Jensen's function* for the function $f(s) = f(\sigma + it)$. Since $\varphi(\sigma; \gamma, \delta)$ with fixed $\gamma$ and $\delta$ is continuous as a function of $\sigma$, it follows from the theorem that $\varphi(\sigma)$ is continuous.

THEOREM 3. *Let* $f_1(s)$, $f_2(s)$, ... *be a sequence of a.-p. functions in the strip* $(\alpha, \beta)$, *which converges uniformly in this strip to the function* $f_0(s)$. *If for all* $n = 0, 1, 2, \ldots$ *the functions* $f_n(s)$ *differ from identical zero, Jensen's function* $\varphi_n(\sigma)$ *converges uniformly as* $n \to \infty$ *to* $\varphi_0(s)$ *(Jensen's function for* $f_0(s)$*) in the interval* $(\alpha, \beta)$. *In this sense Jensen's function* $\varphi(\sigma)$ *depends continuously on the function* $f(s)$.

THEOREM 4. *Let* $f(s)$ *be an a.-p. function in the strip* $(\alpha, \beta)$, *not being identically zero. Its corresponding Jensen function* $\varphi(s)$ *is convex in* $(\alpha, \beta)$ *and for any* $\sigma \in (\alpha, \beta)$ *the four mean movements satisfy the inequality*

$$\varphi'(\sigma - 0) \leqslant \underline{C}^{\pm}(\sigma) \leqslant \overline{C}^{\pm}(\sigma) \leqslant \varphi'(\sigma + 0).$$

*Also two densities of zeros in each strip* $(\sigma_1, \sigma_2)$ $(\alpha < \sigma_1 < \sigma_2 < \beta)$ *satisfy the inequalities*

$$\frac{1}{2\pi}[\varphi'(\sigma_2 - 0) - \varphi'(\sigma_1 + 0)] \leqslant \underline{H}(\sigma_1, \sigma_2) \leqslant \overline{H}(\sigma_1, \sigma_2) \leqslant \frac{1}{2\pi}[\varphi'(\sigma_2 + 0) - \varphi'(\sigma_1 - 0)].$$

From theorem 4 we can obtain three corollaries.

COROLLARY 1. If the function $\varphi(\sigma)$ is differentiable at a point $\sigma$, the left-hand and right-hand mean movements $C^-(\sigma)$ and $C^+(\sigma)$

of the function $f(\sigma + it)$ exist and have the common value

$$C^-(\sigma) = C^+(\sigma) = \varphi'(\sigma).$$

COROLLARY 2. If $\varphi(\sigma)$ is differentiable at two points $\sigma_1$ and $\sigma_2$, there exists a density of zeros of $f(s)$, $H(\sigma_1, \sigma_2)$ in the strip $(\sigma_1, \sigma_2)$ where

$$H(\sigma_1, \sigma_2) = \frac{1}{2\pi}[\varphi'(\sigma_2) - \varphi'(\sigma_1)].$$

The last formula, on an analogy with the classical formula of the theory of analytic functions, is naturally called *Jensen's formula.*

COROLLARY 3. The following formula is valid

$$\frac{1}{2\pi}[\varphi'(\sigma + 0) - \varphi'(\sigma - 0)] = \lim_{\varepsilon \to 0} \overline{H}(\sigma - \varepsilon, \sigma + \varepsilon).$$

From this formula it follows that $\varphi(\sigma)$ is differentiable at those points and only those where

$$\lim_{\varepsilon \to 0} \overline{H}(\sigma - \varepsilon, \sigma + \varepsilon) = 0.$$

# REFERENCES

## CHAPTER I


1. P. S. ALEKSANDROV, *Introduction to the general theory of sets and functions* (*Vvedenie v obshchuyu teoriyu mnozhestv i funktsii*) I, GITTL, 1948.
2. P. S. ALEKSANDROV and A. N. KOLMOGOROV, *Introduction to the theory of functions of a real variable* (*Vvedenie v teoriyu funktsii deistvitel'nogo peremennogo*), GONTI, 1938.
3. S. KACZMARZ and H. STEINHAUS, *Theorie der Orthogonalreihen*, Monogr. Mat. Bd. 6, Warsaw and Lvov 1935.
4. A. N. KOLMOGOROV and S. V. FOMIN, *Elements of the theory of functions and functional analysis* (*Elementy teorii funktsii i funktsional'nogo analiza*), Izd. MGU, I, 1956 and II, 1959. (Engl. transl.: Greylock Press, Rochester, N.Y., 1957.)
5. H. LEBESGUE, Leçons sur l'integration et la recherche des fonctions primitives, *Collections de monographies sur la theorie des fonctions*, Paris 1950.
6. N. N. LUZIN, *The theory of functions of a real variable* (*Teoriya funktsii deistvitel'nogo peremennogo*), Uchpedgiz, 1940.
7. I. P. NATANSON, *The theory of functions of a real variable* (*Teoriya funktsii veshchestvennogo peremennogo*), GITTL, 1957.
8. S. SAKS, *Theory of the integral*, 2nd rev. edit. (Eng. trans.: Stechert, New York 1937).


## CHAPTER II


1. N. I. AKHIEZER, *Lectures on the theory of approximation* (*Lektsii po teorii approksimatsii*), M. 1947. (Engl. transl.: Ungar, New York 1956).
2. N. I. AKHIEZER, On the weighted approximation of continuous functions by polynomials on the whole axis (O vzveshennom priblizhenii nepreryvnykh funktsii mnogochlenami na vsei chislovoi osi), *UMN* **11**, 4 (70), 3–43 (1956).
3. S. N. BERNSTEIN, *Collected works* (*Sobranie sochinenii*) I, 1952, and II, 1954.
4. S. N. BERNSTEIN, *Extremal properties of polynomials and the best approximation of continuous functions of one real variable* (*Extremal'nye svoistva polinomov i nailuchshee priblizhenie nepreryonykh funktsii odnoi veshchestvennoi peremennoi*), ONTI, 1937.
5. A. O. GEL'FOND, *The calculus of finite differences* (*Ischislenie konechnykh raznostei*), Gostekhizdat, 1952.
6. A. O. GEL'FOND, On uniform approximations by polynomials with integral rational coefficients (O ravnomernykh priblizheniyakh mnogoch-

lenami c tselymi ratsional'nymi koeffitsientami), *UMN*, **10**, 1 (63), 41–65, 199–200 (1955).
7. V. L. Goncharov, *The theory of interpolation and approximation of functions* (*Teoriya interpolirovaniya i priblizheniya funktsii*), Gostekhizdat, 1954.
8. S. M. Lozinskii and I. P. Natanson, The metric and constructive theory of functions of a real variable (Metricheskaya i konstruktivnaya teoriya funktsii veshchestvennoi peremennoi), Sb. *Matematika v SSSR za sorok let* 1917–1957, **1**, 295–380, Fizmatgiz (1959).
9. L. A. Lyusternik, On the calculation of values of functions of one variable, *Matematicheskoe prosveshcheniya*, Nos. 3–4, 1959.
10. S. N. Mergelyan, Weighted approximation by polynomials (Vesovye priblizheniya mnogochlenami), *UMN*, **11**, 5 (71), 107–152 (1956).
11. I. P. Natanson, *The constructive theory of functions* (*Konstruktivnaya teoriya funktsii*), Gostekhizdat, 1949.
12. S. M. Nikol'skii, The approximation of functions of a real variable by means of polynomials (Priblizhenie mnogochlenami funktsii deistvitel'nogo peremennogo), Sb. *Matematika v SSSR za tridtsat let* 1917–1947, M–L, 288–318 (1948).
13. S. M. Nikol'skii, The approximation of periodic functions by means of trigonometric polynomials (Priblizhenie periodicheskikh funktsii trigonometricheskimi mnogochlenami), *Trudy Matem. in-ta AN SSSR*, **15**, 1–76 (1945).
14. S. M. Nikol'skii, Inequalities for integral functions of finite degree and their application in the theory of differentiable functions of many variables (Neravenstva dlya tselykh funktsii konechnoi stepeni i ikh primenenie b teorii differentsiruemykh funktsii mnogikh peremennykh), *Trudy Matem. in-ta AN SSSR*, **38**, 244–278 (1951).
15. E. Ya. Remez, *General numerical methods of Chebyshev approximation* (*Obshchie vychislitel'nye metody chebyshevskogo priblizheniya*), Kiev 1957.
16. J. F. Steffensen, *Interpolation*, 2nd ed., Chelsea, N. Y. 1950.
17. S. B. Stechkin, On the order of best approximations of continuous functions (O poryadke nailuchshikh priblizhenii nepreryvnykh funktsii), *IAN SSSR*, **15**, 219–242 (1951).
18. A. S. Householder, *The principles of numerical analysis* McGraw-Hill, New York, 1953.
19. A. F. Timan, *The theory of approximation of functions of a real variable* (*Teoriya priblizhenii funktsii deistvitel'nogo peremennogo*), Fizmatgiz, 1960. (Eng. trans.: Pergamon Press, 1963.)
20. I. S. Berezin and N. P. Zhidkov, *Computing methods* (*Metody vychislennii*), I and II, Fizmatgiz, 1959. (Eng. trans.: Pergamon Press 1965.)

## CHAPTER III

1. H. Bohr, *Almost-periodic functions* (Chelsea), New York 1951.
2. B. M. Levitan, *Almost-periodic functions* (*Pochti-periodicheskie funktsii*), Gostekhizdat, 1953.

# INDEX OF SYMBOLS


$[a, b]$ — the segment $a \leqq x \leqq b$ 1
$(a, b)$ — the interval $a < x < b$ 1
$B$-measurable sets — measurable in the Borel sense 18
$B^p$ — Besicovitch space 192
$C$ — the space of continuous functions 80
$CP_{[0,2\pi]}$ — space of continuous periodic functions of period $2\pi$ 80
$A \cap B$ — the intersection of the sets $A$ and $B$ 8
$A \cup B$ — the union of the sets $A$ and $B$ 8
$\bigcup_a E_a$ (or $\sum_a E_a$) — the union of the sets $E_a$ 8
$\bigcap_a E_a$ (or $\Pi_a E_a$) — the intersection of the sets $E_a$ 8
$A \setminus B$ — the difference of the sets $A$ and $B$ 8
$x \in M$ — the element $x$ belongs to the set $M$ 1
$x \bar{\in} M$ — the element $x$ does not belong to the set $M$ 1
$A \subseteq B$ (or $A \subset B$) — the set $A$ is a subset of set $B$ 2
$\varnothing$ — the empty set 2
$A \sim B$ — the sets $A$ and $B$ are equivalent 4
$\aleph_0$ ("aleph-zero") — the power of a denumerable set 4
$\mathfrak{C}$ and $\aleph$ ("aleph") — the power of the continuum 7
$E'$ — the derived set of the set $E$ 11
$\overline{E}$ — the closure of the set $E$ 11
lim sup $E$ (or $\overline{\lim} E$) — the upper limit of the set $E$ 12
lim inf $E$ (or $\underline{\lim} E$) — the lower limit of the set $E$ 12
lim sup $a_n$ (or $\overline{\lim_{n\to\infty}} a_n$) — the upper limit of the sequence of numbers $a_n$ 12
lim inf $a_n$ (or $\underline{\lim_{n\to\infty}} a_n$) — the lower limit of the sequence of numbers $a_n$ 12
$mE$ — the measure of the set $E$ 16
$m_e E$ ($m_i E$) — the outer (inner) measure of the set $E$ 17
lim sup $E_n$ (or $\overline{\lim_{n\to\infty}} E_n$) — the upper limit of the sequence of sets $E_n$ 20
lim inf $E_n$ (or $\underline{\lim_{n\to\infty}} E_n$) — the lower limit of the sequence of sets $E_n$ 20
$\bigvee_a^b$ (or $\bigvee$ ) — the total variation of a function on the segment $[a, b]$ 26
$E\{x : f(x) > a\}$ or $E\{f > a\}$ — the set of points $x$ which satisfy the condition $f(x) > a$ 29
$D_s f(x_0)$, $D_d f(x_0)$ — the left-hand and right-hand derivatives of the function $f$ at the point $x_0$ 33
$f^+(x_0)$, $f^-(x_0)$ — the upper and lower derivatives of the function $f$ 33, 34

$\lambda^+$. $\lambda^-$, $\Lambda^+$, $\Lambda^-$ — the Dini derivative numbers 34, 35
$f^{[1]}$ — the asymptotic (approximative) derivative of the function $f$ 35
$\bar{I}$, $\underline{I}$ — the upper and lower Darboux integrals 37
$\delta_{m,n}$ — Kronecker's symbol 60
$\mathrm{Lip}_M 1$ — a Lipschitz condition with constant $M$ 129
$S_n(x; f)$ — the $n$-th Fourier sum 133
$m_e(E)$, $m_i(E)$ — the outer and inner $n$-dimensional Lebesgue measure 72
$v_k(E)$ — the $k$-dimensional Hausdorff measure 73
$L$ — the space of summable functions 81
$L^p$ — the space of functions which are summable in the *pth* degree 82
$M$ — the space of essentially bounded functions 83
$V$ — the space of functions of bounded variation 83
$S$ — the space of measurable functions 84
$l^p$ — the space of sequences 84
$V_{n,p}$ — the de la Vallée-Poussin sum 137
$J_n(x; f)$ — the de la Vallée-Poussin integral 137
$\sigma_n(x; f)$ — the Féjer sum 135
$D_n(u)$ — the Dirichlet kernel 133
$H$ — Hilbert space 159
$\|f\|$ — the norm of the function $f$ 143
$\varepsilon, N$-almost-period 184
$S^p$ — Stepanov space 188
$\varepsilon$-almost-period 189
$W^p$ — Weyl space 190, 191
$B^p$-distance 192
$C$-property 31
$N$-property 29
$S$-distance of order $p$ ($S^p$-distance) 188
$W$-distance of order $p$ 190
$\varepsilon$-neighbourhood of a point 10, 70
$\varepsilon$-almost-period ($\varepsilon$-displacement) 170, 189

# INDEX


Absolute value of a solution 187
Aggregate 1
Argument, left-hand 201
Axioms of the metric 78

Basis 178
Best approximation in the mean 62
Best method of approximation of functions with respect to a system $\Phi$ 144
Best uniform approximation 111

Cauchy's criterion 79
Closure of a set 11
Complement 8
Component of connectivity 75
Conditions
  Dini-Lipschitz 134
  Lipschitz 27, 33
Constructive theory of functions 168
Continuity of the Lebesgue integral, absolute 49
Continuum hypothesis 14
Convergence in norm (strong) 79, 80
Covering of a set 17
Cut of a function 43

Density of zeros 203
Derivative 31
  asymptotic (approximate) 35
  generalized 148
    in the Sobolev sense 50
    mixed partial 155
    partial 155
  left-hand (to the left) and right-hand (to the right) 33
Derivative
  partial 74
  point 32
  symmetric (Schwartzian) 36
  upper (lower) 33
Derivative number 34
  Dini 34, 35
  upper (lower) right-hand (left-hand) 34, 35
Determinant, Gram 160
Difference between sets 8
Difference ratio 31
Differential
  asymptotic 74
  exact total 74
Distance 81, 188
  between elements 78
  of a function from a generalized polynomial 163
  of an element from a set 158
Domain 13, 71
  of definition of a function 22

Element
  null 80, 81
  of a sequence 2
  of a set 1
Entropy 161

Family 1
  $S^p$-compact 189
Formula
  for integration by parts 50
  Jensen's 205
  Newton–Leibniz 47
Fourier
  coefficients 61, 140, 147, 174
  expansion (series) 61
  exponents 174

Function 22
absolutely continuous 28, 29, 47
almost-periodic 190
in the Bochner sense 171
in a strip 195
on a group 194
analytic 127
asymptotically
continuous 35
differentiable 35
Besicovitch 192
$B^p$-almost-periodic 192
bounded 23
characteristic 30, 67
continuous
at a point with respect to a set 23
with respect to a set 74
differentiable 31, 74
Dirichlet 24
$\varepsilon, N$-almost-periodic 184
essentially bounded 83
finite 23
harmonic almost-periodic 200
increasing (non-decreasing) and decreasing (non-increasing) 25
integral 96, 122
integrating (weighted) 38
of exponential type 129, 153
Jensen's 203
jump 25
Lebesgue integrable (summable) 41, 42, 43, 45, 52
limit 52
limited periodic 183
measurable 53, 72
on a set 29
with respect to a measure $\mu$ 78
monotonic 25
of a point (variables) 71
of a set 71
additive 71
completely additive 71
of bounded variation 26, 33, 75
Rademacher 65
Riemann integrable 37
Function
Stepanov almost-periodic 188
strictly decreasing (increasing) 25
strictly monotonic 25
summable 76, 132
$S^p$-almost-periodic 189
$S^p$-normal 189
$S^p$-uniformly continuous 189
uniform almost-periodic 170
uniformly continuous 23
weight 66, 100, 131
Weyl 190

Group
abstract 193
commutative (Abelian) 194

Hermite's problem 95

Image of a set 23
Inequality
(axiom), triangle 79
Bernstein's 121
Bessel's 63, 173, 174
generalized 154
Landau's 140
Integral
Darboux 37
Lebesgue 40, 43, 50, 76
indefinite 47
Lebesgue–Stieltjes 78
Riemann 37, 41
Stieltjes (Riemann–Stieltjes) 38
de la Vallée-Poussin singular 137
Intersection 8
Interval 1
adjoint 14
complementary 14
Inverse image of a point 23

Jump of a function 24, 39, 40

Kernel
Dirichlet 133
Féjer 136
Kronecker's symbol 60

Limit
of a sequence 11
upper (lower) 12
of sets, upper (lower) 12
right-hand (on the right) and left-hand (on the left) of a function 24
passage to, under the integral sign 54
Locus 1

Matrix, normal 101
Measure 16
Hausdorff
linear 73
$k$-dimensional 73
plane (two-dimensional) 73
inner 17
Lebesgue 17
$n$-dimensional (and inner) 72
$n$-dimensional of a measurable set 72
($n$-dimensional volume) 71
outer 17
Modulus of continuity 119
Movement, mean (upper and lower, left-hand and right-hand) 202
mean of an almost-periodic function 184
Multiplicativity 65

Nearness, uniform and mean degree 107, 108
Neighbourhood of a point 10
Norm 79, 81, 82
of a function 146
uniform 188
Number, rational 4

One-to-one correspondence 4
Orthogonalization 68

Parseval's equation 175, 176, 177, 191
Point
angular 33
condensation 14
density 19
inner 24
isolated 10
limit 10
of convergence of a sequence of functions 52
of discontihuity of a function 24
of $n$-dimensional space 70
of rarefication 19
Points
Chebyshev 99
multiple 95
simple 95
Polynomial, best approximating 111
Polynomials
Abel–Goncharov 106
Bernstein 105, 109
Chebyshev 89, 114, 141, 142
fundamental 97
generalized 163
Hermite 95, 148
interpolation 88
generalized 104
Jacobian 147
Lagrange 89
Laguerre 148
Legendre 65, 141
Newton's 91
real trigonometric 153
trigonometric 92
Bochner–Féjer 179
Power 4
of the continuum 7
Product 8
Proper part 2

Range of values of a function 23
Recurrence relations 3

Scalar product 82
Schmidt's process of orthogonalization 69
Section of a set 73

Segment 1
Sequence 1
  convergent 52
    almost everywhere 52
    at a point 79
    in measure 57
    in the mean 56
    uniformly 53
  fundamental 79
  of sets
    convergent 21
    divergent 21
    monotonically decreasing (increasing) 21
Series
  convergent 52
  Dirichlet 197
  Fourier 61, 133, 141, 147
  Fourier–Riemann (Fourier–Lebesgue, Fourier–Stieltjes) 62
  functional 53
  orthogonal 61
Set 1
  Borel 18
  bounded 2
  Cantor perfect 15, 48
  Chebyshev 92, 164
  closed 11, 75
  conditionally compact 171
  connected 75
  dense in itself 15
  dense on a set $R$ 15
  denumerable 4
  derived 11
  empty 2
  everywhere dense 15
  linearly independent 178
  measurable $B$ ($B$-set) 18
  measurable in the Lebesgue sense 17
  non-denumerable 14
  nowhere dense 15
  of complete measure 19
  of level 75
Set
  of the first category 16
  of the second category 16
  open 13
Set
  perfect 14
point 1
  reduced 18
  relatively dense 170
Sets
  equivalent 4
  of equal power 4
Space
  ($B^p$) Besicovitch 192
  complete 79
  ($C$) of continuous functions 80
  ($CP_{[0,2\pi]}$) of continuous periodic functions 80
  Euclidean of $n$ dimensions 70
  function 79
  ($H$) Hilbert 147, 158
    ($L$) 81
  linear 79
    normed 146, 154
  ($L^p$) 82
  ($l^2$ and $l^p$) of sequences 83
  ($L^2$) of functions with a summable square 81
  metric 78
  ($M$) of essentially bounded functions 83
  normed 79
  strictly normed 158
  ($S$) of measurable functions 84
  ($S^p$) Stepanov 188
  ($V$) of functions of bounded variation 83
  ($W^p$) Weyl 191
Sphere 79
  closed (and open) 70, 71
Strip 195
Subset 1
  proper 2
Sum
  Darboux 36
  Darboux–Stieltjes 38
  Féjer 135
  Fourier 132
  Lebesgue 40, 76
  of sets 7
  partial 53
  de la Vallée-Poussin 137
Superposition of functions 29

System of functions
Chebyshevian 92
closed 61, 160
complete 61
fundamental 104
linearly independent 68
normalized 60
orthogonal 60, 69
orthogonal with weight 66
orthonormal 60
trigonometrical 64
Walsh 66
Systems of functions equivalent 68

Term of a sequence 2
general 2
Theorem
approximation 178, 179, 186, 197, 201
Bochner 171
Bohl–Bohr 172
Bohr 177, 181, 184
Bolzano–Weierstrass 13
Borel 30
Borel–Lebesgue 22
Cantor–Bendixson 14
Cantor–Bernstein 6
Darboux 32
Denjoy 35, 36
Dini 35
Egorov 56
Fichtengol'ts 29
Fubini 77
Gram 68
Kolmogorov 64
Kolmogorov–Seliverstov–Plessner 64

Theorem
Kronecker 182
Kronrod 75
Lebesgue 37, 57
Luzin 31, 35, 56
mean value 172, 194
Menshov–Rademacher 64
Merkh'yuber 164
of absolute continuity of the Lebesgue integral 46, 47
of complete additivity of the integral 41, 44
of finite additivity of the Lebesgue integral 45
of the mean 41
Rademacher 64
Riemann 37
Riezs–Fischer 63
Schur 69
Stepanov 74
uniqueness 177, 186, 197, 201
Vitali 22
Weierstrass 24, 56

Union of sets 7
Upper (lower) bound 11
Upper bound, essential 83
point 11

Variation
linear 75
of a function of Segment 36
plane 75
total 26, 49

# OTHER TITLES IN THE SERIES IN PURE AND APPLIED MATHEMATICS

Vol. 1. WALLACE — *An Introduction to Algebraic Topology*
Vol. 2. PEDOE — *Circles*
Vol. 3. SPAIN — *Analytical Conics*
Vol. 4. MIKHLIN — *Integral Equations*
Vol. 5. EGGLESTON — *Problems in Euclidean Space: Application of Convexity*
Vol. 6. WALLACE — *Homology Theory on Algebraic Varieties*
Vol. 7. NOBLE — *Methods Based on the Wiener–Hopf Technique for the Solution of Partial Differential Equations*
Vol. 8. MIKUSIŃSKI — *Operational Calculus*
Vol. 9. HEINE — *Group Theory in Quantum Mechanics*
Vol. 10. BLAND — *The Theory of Linear Viscoelasticity*
Vol. 11. KURTH — *Axiomatics of Classical Statistical Mechanics*
Vol. 12. FUCHS — *Abelian Groups*
Vol. 13. KURATOWSKI — *Introduction to Set Theory and Topology*
Vol. 14. SPAIN — *Analytical Quadrics*
Vol. 15. HARTMAN and MIKUSIŃSKI — *Theory of Lebesgue Measure and Integration*
Vol. 16. KULCZYCKI — *Non-Euclidean Geometry*
Vol. 17. KURATOWSKI — *Introduction to Calculus*
Vol. 18. GERONIMUS — *Polynomials Orthogonal on a Circle and Interval*
Vol. 19. ELSGOLC — *Calculus of Variations*
Vol. 20. ALEXITS — *Convergence Problems of Orthogonal Series*
Vol. 21. FUCHS and LEVIN — *Functions of a Complex Variable, Volume II*
Vol. 22. GOODSTEIN — *Fundamental Concepts of Mathematics*
Vol. 23. KEENE — *Abstracts Sets and Finite Ordinals*
Vol. 24 DITKIN and PRUDNIKOV — *Operational Calculus in Two Variables and its Applications*
Vol. 25. VEKUA — *Generalized Analytic Functions*
Vol. 26. FASS and AMIR-MOÉZ — *Elements of Linear Spaces*
Vol. 27. GRADSHTEIN — *Direct and Converse Theorems*
Vol. 28. FUCHS — *Partially Ordered Algebraic Systems*
Vol. 29. POSTNIKOV — *Foundations of Galois Theory*
Vol. 30. BERMANT — *A Course of Mathematical Analysis, Part II*
Vol. 31. ŁUKASIEWICZ — *Elements of Mathematical Logic*
Vol. 32. VULIKH — *Introduction to Functional Analysis for Scientists and Technologists*
Vol. 33. PEDOE — *An Introduction to Projective Geometry*
Vol. 34. TIMAN — *Theory of Approximation of Functions of a Real Variable*
Vol. 35. CSÁSZÁR — *Foundations of General Topology*

Vol. 36. BRONSHTEIN and SEMENDYAYEV – *A Guide-Book to Mathematics for Technologists and Engineers*
Vol. 37. MOSTOWSKI and STARK – *Introduction to Higher Algebra*
Vol. 38. GODDARD – *Mathematical Techniques of Operational Research*
Vol. 39. TIKHONOV and SAMARSKII – *Equations of Mathematical Physics*
Vol. 40. MCLEOD – *Introduction to Fluid Dynamics*
Vol. 41. MOISIL – *The Algebraic Theory of Switching Circuits*
Vol. 42. OTTO – *Nomography*
Vol. 43. RANKIN – *An Introduction to Mathematical Analysis*
Vol. 44. BERMANT – *A Course of Mathematicvl Analysis, Part I*
Vol. 45. KRASNOSEL'SKII – *Topological Methods in the Theory of Nonlinear Integral Equations*
Vol. 46. KANTOROVICH and AKILOV – *Functional Analysis in Normed Spaces*
Vol. 47. JONES – *The Theory of Electromagnetism*
Vol. 48. FEJES TÓTH – *Regular Figures*
Vol. 49. YANO – *Differential Geometry on Complex and Almost Complex Spaces*
Vol. 50 MIKHLIN – *Variational Methods in Mathematical Physics*
Vol. 51. FUCHS and SHABAT – *Functions of a Complex Variable and Some of their Applications, Volume I*
Vol. 52. BUDAK, SAMARSKII and TIKHONOV – *A Collection of Problems on Mathematical Physics*
Vol. 53. GILES – *Mathematical Foundations of Thermodynamics*
Vol. 54. SAUL'YEV – *Integration of Equations of Parabolic Type by the Method of Nets*
Vol. 55 PONTRYAGIN *et al.* – *The Mathematical Theory of Optimal Processes*
Vol. 56. SOBOLEV – *Partial Differential Equations of Mathematical Physics*
Vol. 57. SMIRNOV – *A Course of Higher Mathematics, Volume I*
Vol. 58. SMIRNOV – *A Course of Higher Mathematics, Volume II*
Vol. 59. SMIRNOV – *A Course of Higher Mathematics, Volume III, Part 1*
Vol. 60. SMIRNOV – *A Course of Higher Mathematics, Volume III, Part 2*
Vol. 61. SMIRNOV – *A Course of Higher Mathematics, Volume IV*
Vol. 62. SMIRNOV – *A Course of Higher Mathematics, Volume V*
Vol. 63. NAIMARK – *Linear Representations of the Lorentz Group*
Vol. 64. BERMAN – *A Collection of Problems on a Course of Mathematical Analysis*
Vol. 65. MESHCHERSKII – *A Collection of Problems of Mechanics*
Vol. 66. ARSCOTT – *Periodic Differential Equations*
Vol. 67. SANSONE and CONTI – *Non-linear Differential Equations*
Vol. 68. VOLKOVYSKII, LUNTS and ARAMANOVICH – *A Collection of Problems on Complex Analysis*
Vol. 69. LYUSTERNIK and YANPOL'SKII – *Mathematical Analysis–Functions, Limits, Series, Continued Fractions*
Vol. 70. KUROSH – *Lectures in General Algebra*
Vol. 71. BASTON – *Some Properties of Polyhedra in Euclidean Space*
Vol. 72. FIKHTENGOL'TS – *The Fundamentals of Mathematical Analysis, Volume 1*
Vol. 73. FIKHTENGOL'TS – *The Fundamentals of Mathematical Analysis, Volume 2*
Vol. 74. PREISENDORFER – *Radiative Transfer on Discrete Spaces*

Vol. 75. FADDEYEV and SOMINSKII – *Elementary Algebra*
Vol. 76. LYUSTERNIK, CHERVONENKIS and YANPOL'SKII – *Handbook for Computing Elementary Functions*
Vol. 77. SHILOV – *Mathematical Analysis – A Special Course*
Vol. 78. DITKIN and PRUDNIKOV – *Integral Transforms and Operational Calculus*
Vol. 79. POLOZHII – *The Method of Summary Representation for Numerical Solution of Problems of Mathematical Physics*
Vol. 80. MISHINA and PROSKURYAKOV – *Higher Algebra–Linear Algebra, Polynomials, General Algebra*
Vol. 81. ARAMANOVICH *et al.* – *Mathematical Analysis –Differentation and Integration*
Vol. 82. REDEL – *The Theory of Finitely Generated Commutative Semigroups*
Vol. 84. MIKHLIN – *Multidimensional Singular Integrals and Integral Equations*
Vol. 84. LEBEDEV, SKAL'SKAYA and UFLYAND–*Problems in Mathematical Physics*
Vol. 85. GAKHOV – *Boundary Value Problems*
Vol. 86. PHILLIPS – *Some Topics in Complex Analysis*
Vol. 87. SHREIDER – *The Monte Carlo Method*
Vol. 88. POGORZELSKI – *Integral Equations and their Applications, Vol. I, Parts 1, 2 and 3*
Vol. 89. SVESHNIKOV – *Applied Methods of the Theory of Random Functions*

MADE IN GREAT BRITAIN